The Dissertation Committee for Brittany Nicole Hall-Clark Certifies that this is the approved version of the following dissertation:

The Relation of Stereotype Threat to African American and Latino Performance on the WAIS-IV: An Intelligence Malleability Intervention Approach

Committee:

Manuel Ramirez III, Supervisor

Germine Awad

Rebecca Bigler

Kevin Cokley

Michael Telch

The Relation of Stereotype Threat to African American and Latino Performance on the WAIS-IV: An Intelligence Malleability Intervention Approach

by

Brittany Nicole Hall-Clark, B.A., M. A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2011

Dedication

I dedicate this dissertation to the racial/ethnic minority individuals who have been misclassified by intelligence tests and overrepresented in the special education system.

Acknowledgements

I thank my mentor, Dr. Manuel Ramirez, for being committed to the advancement of ethnic minority graduate students, and for helping me to become the first African American to graduate from the clinical psychology program of the University of Texas of Austin. I would also like to thank Drs. Kevin Cokley and Gigi Awad for their guidance, support, and faith in me. I also owe thanks to my fellow graduate students for keeping me accountable, focused, and encouraged. I especially appreciate Anushka for being my "stats coach" and Nanci and Collette for taking the time to edit my work. I pay homage to my late mother whose memory inspires me to be as successful as she was. I thank all of my family members, especially my father, stepmother, in-laws, and sister, for always believing in me. Lastly, I thank my husband, Carl, for his unwavering support and confidence in me.

The Relation of Stereotype Threat to African American and Latino Performance on the WAIS-IV: An Intelligence Malleability Intervention Approach

Brittany Nicole Hall-Clark, Ph.D.

The University of Texas at Austin, 2011

Supervisor: Manuel Ramirez III

Stereotype threat is defined as a sociopsychological threat evoked by an evaluative situation in which a negative stereotype about one's group could be confirmed (Steele, 1997). While the deleterious effects of stereotype threat have been demonstrated numerous times in laboratory settings (McKay, Doverspike, Bowen-Hilton, & Martin, 2002; Ngyuen & Ryan, 2008; Spencer, Steele & Quinn, 1999; Steele & Aronson, 1995), generalization to actual testing situations has been limited (Stricker & Ward, 2004). The current study sought to increase ecological validity by examining stereotype threat among racial/ethnic minority students undergoing assessment using the Wechsler Adult Intelligence Scale-IV (WAIS-IV) without explicit priming. Another aim was to reduce stereotype threat by emphasizing the malleability of intelligence, as recommended by previous researchers (Aronson, Fried, & Good, 2002; Good, Aronson, & Inzlicht, 2003). Additionally, the relationship of ethnic identity to stereotype threat and test performance, and the role of anxiety, a proposed mechanism of stereotype threat, were examined.

Participants were also interviewed about their experiences of stereotype threat using a phenomenological approach.

A 2(condition) x 3(race/ethnicity) experimental design was used, and 138 college students were randomized to the control or malleability conditions. Due to manipulation failure, the hypothesis that African and Latino American students would experience less stereotype threat and perform better on the WAIS-IV in the malleability condition could not be tested. Qualitative findings suggested that while participants endorsed perceptions of stereotype threat in general societal settings, they did not report stereotype threat while undergoing the WAIS-IV. The hypothesis that ethnic identity moderates the relationship between stereotype threat and performance received mixed support: ethnic identity-affirmation interacted with perceived stereotype threat on Digit Span, but all other interactions were nonsignificant. Lastly, the hypothesis that anxiety mediates the relationship between perceived stereotype threat and WAIS-IV performance was not supported. However, post-hoc analyses suggested that perceived stereotype threat mediates the relationship of anxiety and WAIS-IV performance. Correlational results revealed that perceived stereotype threat and stereotype vulnerability were related to WAIS-IV scores. In addition, students of color reported greater test and state anxiety than their European American counterparts. Implications for researchers, test administrators, and admissions officers are discussed.

Table of Contents

INTRODUCTION

Certain racial/ethnic minority groups have consistently performed more poorly on tests of cognitive ability than their European American counterparts (Helms, 1992; Kaufman, McLean & Reynolds, 1988; Yeung & Pfeiffer, 2009). The average European American IQ score is 100, while the average African American score is 85 and the average Latino score is 89 (Herrnstein & Murray, 1994; Nisbett, 2009). Studies conducted after 1980, which tend to control for SES differences, have reported that the gap between African Americans and European Americans has decreased over time, with values ranging from 1 to 9.5 IQ points (Nisbett, 2009; Vincent, 1991). However, it is important to note that commonly used intelligence tests have been reported to label approximately one out of every six African American and Latino American examinees as mentally retarded (Mackintosh, 1998).

Intelligence test scores have important implications for academic and professional placement (McFarland, Lev-Arey, & Ziegert, 2003; Ployhart, Ziegert & McFarland, 2003; Wicherts, Dolan, & Hessen, 2005). If other factors aside from cognitive ability influence the outcomes of these measures, then the validity of such tests should be questioned. Several researchers have noted that underperformance may occur due to reasons other than true ability, such as interpersonal factors of a testing situation (Miller-Jones, 1989), cultural differences (Nisbett, 2009; Ramirez & Gonzalez, 1972), and anxiety (Reeve & Bonaccio, 2008).

One factor thought to contribute to test scores differences is a phenomenon known as stereotype threat, which is posited to depress the performance of racial/ethnic minority

students. Since the introduction of the concept (Steele & Aronson, 1995), effects of stereotype threat have been widely replicated and extended to many social groups. Ryan and Ryan (2005) argued that stereotype threat calls into question the validity of test score interpretation for stereotype threat vulnerable groups. The situation of having one's intelligence evaluated is anxiety-provoking for many, but may be especially so for groups who are stereotyped to be less intelligent. Most of the stereotype threat literature has focused on academic performance and not on intellectual assessments, despite the fact that the central tenet of stereotype threat that negative stereotypes about intelligence interfere with test performance. Considering that the Wechsler Adult Intelligence Scale IV (WAIS-IV) is one of the most commonly used intelligence tests in the field of psychology, examining the potential role of stereotype threat in performance is important. To this end, the current study is the first to examine the relationship of stereotype threat to the WAIS-IV. Another aim of the present study was to address the empirical gap between the many successful laboratory replications of stereotype threat and the relative dearth of findings in real-world settings (Davis & Simmons, 2009; Stricker & Ward, 2008). In the current study, efforts were made to increase the ecological validity of stereotype threat by not explicitly inducing stereotype threat, which is nearly always done in the empirical literature but seldom done in the real world.

The terms race/ethnicity and racial/ethnic are used throughout the manuscript to reflect that while the literature often discusses African American and European American test score differences in terms of race, the current study also considers the influence of cultural variables such as ethnic identity and disengagement.

The Rise of Intelligence Testing: From the Binet-Simon to the WAIS-IV

The Wechsler tests have been the most widely used intelligence tests in the field of psychology, which may suggest relative agreement in the usage of this assessment. However, the history of intellectual assessment has been characterized by several unresolved debates regarding the nature, components, and meaning of intelligence (Gottfredson & Saklofske, 2009; Lohman, 1997; McGrew & Flanagan, 1998; Valencia & Suzuki, 2001). The key persons in the intelligence testing movement often differed in their theoretical orientations and methodological approaches, which reflected the larger debates of the field. To fully understand the deep controversy surrounding the intellectual assessment of ethnic minorities, it is important to consider the historical and sociopolitical influences that contributed to the development of intelligence testing (Guthrie, 1976; Lohman, 1997).

Although theories strive to be objective, they may be inevitably influenced by subjective influences, such as the zeitgeist of an era and theorists' personal experiences and beliefs (Lohman, 1997). Intelligence testing originated in the 19th century, during which eugenics and social darwinism were popular (Guthrie, 1976; Lohman, 1997; Murdoch, 2007). Sir Francis Galton is credited with beginning the testing movement in Europe and was greatly influenced by his half-cousin, Charles Darwin. Galton was a eugenicist concerned with the heredity of genius and aimed to understand factors that would increase the quality of the human gene pool (Guthrie, 1976; Murdoch, 2007). Galton drew on the theory of evolution to form his theories on intelligence. It is widely known that Galton believed in racial hierarchies, writing that the "Greeks were as

superior to his English contemporaries in their capacity for genius as his contemporaries were superior to Africans and their American descendants" (Brody, 2000). Galton believed that bright people should be encouraged to procreate, while the spread of inferior genes of the "feebleminded", or mentally retarded, should be prevented. He argued that intelligence testing was a means to determine the quality of one's genes. Galton's ideas were well-received because aristocracy had long determined those who had access to the best resources in European society. Thus, eugenics represented an ostensibly fairer way to redistribute resources based on merit rather than legacy (Lohman, 1997). Galton believed that physical capabilities reflected mental quality and used measures of reaction time, hand strength, and sensory discrimination to study intelligence (Brody, 2000; Guthrie, 1976; Murdoch, 2007).

Charles Spearman also believed that intellectual functioning was represented by performance on sensory discrimination tasks, based his observation that such tasks correlated with the academic performance of South African school children (Brody, 2000; Murdoch, 2007). In 1904, Spearman published one of the most influential papers in the field of intelligence (Brody, 2000). Spearman proposed a two-factor theory of intelligence: a general factor, *g*, and a specific factor, *s*. He posited that *g* was a common intellectual ability that spanned across many types of activities. In contract, *s* referred to intellectual ability in discrete domains. Although Spearman noted that *g* could not account for all aspects on intelligence, he emphasized *g* as a critical factor in his writings. Spearman argued that the best way to measure *g* was to assess performance across a wide range of abilities, averaging their scores. This concept was known as positive manifold

(Goldstein, Scherbaum, & Yusko, 2010). Due to the theory that *g* underlaid all tests of intellectual ability, the intercorrelations between dissimilar tests were thought to reflect *g*.

Binet, a scholar with a strong research interest in testing, disagreed with Spearman's theory of *g*, believing that the theory was too simple to explain a concept as complex as intelligence. He was opposed to the methodology of assessing intelligence through physiological reactivity, and did not agree that simple cognitive processes were associated with intelligence (Brody, 2000). Instead, Binet argued that intelligence should be studied through higher level thinking processes, such as concentration and abstraction that distinguish between adults and children. He reasoned that learning at grade-level involves a general mental ability that is also related to everyday learning (Gottfredson & Saklofske, 2009). He received the opportunity to test his theories when the French government accepted his proposal to develop a diagnostic tool to identify children who were having trouble keeping up with their classmates (Murdoch, 2007; Thorndike, 1997). The advent of universal public education in France flooded classrooms with children of widely divergent abilities. Previously, only children who were seen as promising attended school while less astute children were sent to work at early ages (Murdoch, 2007). Thus, teachers were tasked with educating children who had been in school for years while also socializing new students to the school system.

To aid the overwhelmed school system, Binet and his colleague Simon aimed to separate "normal" children from those that doctors and educators deemed subnormal. They also sought to distinguish those with low ability from those with low motivation and personality issues (Thorndike, 1997). Their work culminated in 1905 with the Binet-Simon test, which is known as the first modern intelligence test. They developed an

algorithm that compared children's biological and mental ages. Using age-based norms, Binet and Simon determined how children performed relative to their peers. A 7-yr-old child performing similarly to 10-year-olds would have a mental age of 10. Despite his faith in empiricism, Binet believed that intelligence was a difficult construct to reliably quantify and thus was uncomfortable with the practice of assigning numbers to intelligence tests for fear that they would be misinterpreted as precise measurements of ability (Brody, 2000; Murdoch, 2007). Binet emphasized that his purpose was to predict academic performance and not to quantify intelligence.

Despite Binet's hesitance about the accuracy of his test, demand for such a tool was high (Thorndike, 1997). Spearman believed that the various subtests of the Binet-Simon could demonstrate positive manifold and thus the existence of *g* (Goldstein, Scherbaum, & Yusko, 2010; Murdoch, 2007). However, Spearman and Binet were wary of one another. Binet did not believe in *g* as the singular definition of general intelligence, reasoning that two individuals could use differing knowledge and skill sets but achieve the same score. Spearman disagreed with Binet but appreciated the practicality of a test well-suited to assess *g* in his view. Despite the theoretical contrasts between Binet and Spearman, their major contributions have often been paired. The Binet-Simon and *g* theory have widely influenced the course of intelligence research and development (Brody, 2000; Murdoch, 2007). Guthrie (1976) noted that that Spearman's views were consistent with the Victorian middle class values of abstraction and verbalization and thus his views were readily accepted. Thus, the eugenicist spirit of intelligence testing continued to spread.

Goddard was the American psychologist who introduced the Binet-Simon to the United States by instituting its use in schools (Murdoch, 1997; Thorndike, 1997). Faced with the challenges of universal public education, the American school system needed a method to sort students of varying abilities (Murdoch, 2007; Thorndike, 1997). Goddard endorsed Galton's hereditarian ideas, although instead of administering intelligence tests, Goddard based his conclusions on behavior, reputation, and medical records, yet did not consider performance on intelligence tests a factor. His well-known book *The Kallikak Family: A Study in the Heredity of Feeble-Mindedness*, contributed to a societal fear of mental retardation (Murdoch, 2007). Despite questionable research methods, the Kallikak book was far-reaching in its influence. "Degeneration," or the reduction of racial purity was a common concern in Europe and America (Lohman, 1997). Degeneration was thought to be caused by hyperfertility in "lower quality races." After American slaves were freed, the genetic consequences of racial intermixing were feared. Lohman (1997) argued that this phobia of degeneration catalyzed the acceptance of intelligence testing as a means to determine genetic quality.

Immigration represented another threat of degeneration; thus, intelligence tests were used to screen prospective immigrants to America on Ellis Island (Guthrie, 1976). Intelligence tests were also used to institutionalize American citizens found to be feebleminded and to legitimize the sterilization of allegedly feebleminded women (Murdoch, 2007). This atmosphere of hereditarianism, xenophobia, and a failure to realize the bias of giving heavily-verbally loaded tests to immigrants contributed to the conclusions that non-Whites were genetically inferior to Whites (Guthrie, 1976; Murdoch, 2007; Thorndike, 1997).

Another American psychologist, Lewis Terman, was influential in the widespread use of intelligence tests in America (Thorndike, 1997). Terman translated and revised the Binet-Simon into the Stanford-Binet, which became a template for American group intelligence tests. Terman believed that 90% of academic problems were due to mental inferiority and believed that all children should be tested to assure they were in classrooms appropriate for their development (Thorndike, 1997). Terman began the convention of multiplying IQ scores by 100 (Guthrie, 1976; Murdoch, 1997). In addition, he developed the first classification system of IQ scores, using terms such as moron, idiot, and imbecile, to assign qualitative descriptions of their projected abilities. The concept of IQ was counter to Binet's dynamic construct of mental growth, considering that children's placements could change every year in school. In contrast, IQ conveyed a stable position, suggesting that scores are not likely to change. William Stern, who created the IQ score in 1912, shared Binet's fear that the impression of precision would render the IQ vulnerable to misinterpretation (Guthrie, 1976). Terman has been criticized for extrapolating beyond his data to claim that mental tests reveal true biological capacity (Thorndike, 1997). Nonetheless, Terman's convention of the IQ score became popular, perpetuating the notion that intelligence tests reveal innate abilities.

Another thrust behind the intelligence testing movement was the expansion of the field of psychology. Intelligence testing was also seen as a way to legitimize psychology as a true science, which was seen as inferior to hard sciences such as biology and physics (Guthrie, 1976; Murdoch, 2007; Thorndike, 1997). Some argue that Robert Yerkes aimed to improve the reputation of psychology by serving the armed forces WWI (Murdoch, 2007). Psychologists during WWI transformed the Stanford-Binet into two group tests for

the American Army. The Alpha test was designed for literate soldiers, and consisted of verbal tests such as arithmetic, vocabulary, and digit memory. In contrast, the Beta was created for illiterate soldiers and included performance tests, such as mazes and picture-oriented problems. Scholars have noted that the army testing environment was one of irreverence, confusion, and disdain regarding intelligence testing (Guthrie, 1976; Murdoch, 2007). There were methodological inconsistencies that should have warranted caution in interpreting test scores. For example, many immigrants were not given the Beta after poor performance on the Alpha, and many African American servicemen were given the Beta, despite being able to read (Guthrie, 1976, Murdoch, 2007). Furthermore, because the Beta was meant to be a nonverbal test, the instructions were given in pantomime, which was likely confusing for examinees (Guthrie, 1976; Murdoch, 2007). Despite these administrative issues, results from the army studies reached the familiar conclusion that immigrants and African Americans were genetically inferior to European Americans (Murdoch, 2007). These data were used to propose bans against interracial marriage, reducing money allocated for African American schools, and to support segregation (Guthrie, 1976).

After WWI, intelligence testing in America boomed, with companies such as the Psychological Corporation (now Pearson Assessment) forming to continue the development of intelligence testing (Gottfredson & Saklofske, 2009). David Wechsler established himself as a front-runner in the intelligence testing field by becoming the first to combine both verbal and performance items into one test. He believed that assessing various types of ability could be useful in understanding cognitive functioning and personality (Kaufman, 2000; Murdoch, 2007). Wechsler created the Wechsler-Bellevue

Intelligence Scale in 1939. Wechsler served as an army test administrator during WWI, which largely contributed to the development of the scale. He drew on the army Alpha and Beta to create the verbal and performance subtests. Similar to Binet and Stern, Wechsler was also hesitant to claim that IQ scores derived from his tests exhaustively measured intelligence. While he was a proponent of the theory of *g*, he believed that *g* alone did not sufficiently capture intelligence (Murdoch, 2007). Wechsler did not include racial/ethnic minorities in his original standardization sample and warned that his test should not be used for these groups (Guthrie, 1976). However, racial/ethnic minorities were frequently tested in spite of Wechsler's caveat. In later versions of the test, racial/ethnic minorities were included in validation procedures (Wechsler, 2008).

The Wechsler Adult Intelligence Scale (WAIS) was introduced in 1955 and has undergone four revisions, with the most recent occurring in 2008. Over time, some subtests have been added, dropped, and revised, but the basic format has been similar (Williams, 1996). While the basic procedures of testing have not changed greatly, their use and interpretation changed more in the 1980s and 1990s than in the previous 50 years (Thorndike, 1997). The Digit Span, Block Design, and Object Assembly subtests originate from screening tests used on Ellis Island. Six of the seven verbal subtests came from the Army Alpha and the Stanford-Binet (Similarities, Comprehension,) while four of seven performance tests (Picture Arrangement, Digit-Symbol Coding, and Picture Completion) came from the Army Beta. Wechsler's tests have enjoyed great success, becoming the most widely used and accepted measure of intelligence in the United States (McGrew & Flanagan, 1998).

It can be argued that the historical context surrounding the development of intelligence tests facilitated the belief that intelligence is a naturally endowed trait. Thus, intelligence tests held wide appeal on the basis that they were a fair way to determine the distribution of educational and financial resources. However, intelligence tests have been criticized for being atheoretical, given that Spearman's theory was applied to Binet's test post-hoc (Gottfredson & Saklofske, 2009; Valencia & Suzuki, 2001). In addition, there has been much debate regarding the definitions and structure of intelligence (Gottfredson & Saklofske, 2009; Lohman, 1997; McGrew, 1998). Accordingly, the use of such tests to determine ability has been questioned. In particular, there has been much concern about the racial/ethnic differences typically observed on intelligence tests, which can ultimately limit access to important resources for these groups (Awad, 2007; Ramirez & Gonzalez, 1972; Valencia & Suzuki, 2001).

Assessment of Racial/Ethnic Minority Populations

The intelligence testing of ethnic minority groups has had a long and controversial history, largely due to the persistent performance discrepancy between different racial/ethnic groups. Further complicating matters is the fact that the concept of race is as murky as the construct of intelligence itself. Although most researchers acknowledge that race is a social construct, there remains debate about the biological differences between the so-called races (Hunt & Carlson, 2007). The biological conceptualization of race distinguishes groups of people by their geographical origins and phenotypical characteristics, such as hair texture, skin color and body type. However, it is commonly known that more genetic variation exists within races than between races (Loehlin, 2000;

Neisser et al., 1996), which draws the biological definition of race into question (for a thorough discussion see Smedley & Smedley, 2005). Furthermore, given the legacy of slavery in America, most African Americans have some proportion of European ancestry. Immigration from around the world has also contributed to much genetic intermixing in the American population. Others argue that despite genetic ambiguity, self-identification serves as a reasonable indictor of race because self-report and genomic markers tend to correlate highly (Hunt & Carlson, 2007). Despite the questionability of race as a scientific concept, it cannot be denied that racial/ethnic group membership has important societal implications for power, privilege, access to resources, and discrimination. Accordingly, the sociopolitical definition of race as a social construction is more widely accepted (Smedley & Smedley, 2005).

The controversy surrounding the intellectual assessment of racial/ethnic minority populations is largely due to the meaning ascribed to test score differences, which take on such importance because of the impact that tests can have in shaping a person's educational and professional opportunities. As with most psychological phenomena, the general conclusion is that both biological and environmental factors contribute to the general factor known as intelligence. Most modern scholars acknowledge that intelligence results from both genetic and environmental factors and their interactive effects. However, the relative contributions of each are still debated. The heritability of intelligence is usually cited to be 60-90% (Ceci, 1996; Gottfredson & Saklofske, 2009). There is a large body of evidence that supports a substantial genetic component of intelligence generated by studies of heterozygous and monozygous twins, adoption, and selective breeding (Hunt & Carlson, 2007; Jensen, 1969; Loehlin, 2000; Neisser et al.,

1996). However, while there is fair agreement of a substantial genetic component of intelligence, researchers debate about what this means.

In addition to a substantial genetic component of intelligence, a myriad of environmental factors, such as en utero conditions, nutrition, SES, hazardous chemicals, economic and educational resources, educational quality, home environment, and cultural values also contribute to the expression of intelligence (Murdoch, 2007; Neisser et al. 1996; Nisbett, 2009). Another testament to the substantial role of environmental influences is the fact that IQ scores have steadily increased over time, despite the restandardization of norms (Murdoch, 2007; Neisser et al., 1996). The most famous observer of this phenomenon was Flynn (1984), who observed that the population mean of IQ tends to increase every decade by three points. In addition, the decreasing test score gap during times of socioeconomic growth and political progress for African Americans reflect the impact of environmental influences and the malleability of intelligence (Hunt & Carlson, 2007).

In response to the controversy surrounding the publication of the inflammatory book *The Bell Curve* (Herrnstein & Murray, 1994), the American Psychological Association (APA) convened a task force to comment on the state of the field (Neisser et al., 1996). The APA Task Force concluded that there is little direct evidence to support the hereditarian point of view that racial/ethnic differences in intelligence are genetic (Neisser et al., 1996). Neisser et al. cautioned against generalizing between individual and group differences, noting that group differences reveal little about a given individual. They concluded that racial/ethnic differences neither resulted from any obvious bias in test construction or administration, nor any one environmental cause. Environmental

explanations thus far proposed to explain test score gaps have accounted for little variance of test scores. This lack of empirical support leaves many questions unresolved. The reasons for group differences remain unclear, although they are within the range of effect sizes produced by environmental factors (Neisser et al., 1996). Racial/ethnic differences are complex, likely resulting from a combination of genetic, environmental and interactional effects (Hunt & Carlson, 2007; Neisser et al., 1996).

Hunt and Carlson (2007) noted that although individually environmental factors explain little variance that the summation of all such effects can contribute to a big effect. Understanding environmental factors contributing to test score differences is especially important, because they are more amenable to change than genetic influences. Thus, research focused on decreasing test score differences is still needed to increase access for historically disenfranchised groups.

In addition to philosophical and theoretical debates about the definition of intelligence and the meaning of racial/ethnic differences, there has been much controversy about the practice of intelligence testing. Intelligence tests have been regarded as the best available predictor of job success (Goldstein et al., 2010; Neisser et al., 1996). Correlations between intelligence tests and school performance, grades, and work performance have consistently been reported to between .3 and .5 (Hunt & Carlson, 2007; Neisser et al., 1996). Some regard this number as an accomplishment, arguing that intelligence tests are one of the most consistent predictors at our disposal. Others are more critical, focusing on the fact that a correlation of .5 translates to 25% of the variance, leaving much unexplained by test scores (Williams, 1996).

In spite of modest predictive validity, intelligence tests continued to be commonly used in American schools from the preschool to graduate levels. Instruments such as the WAIS-IV are used to diagnose mental retardation and learning disabilities and to assess underachievement and giftedness. Psychologists use such tests to inform educational admissions and tracking, as well as vocational counseling and hiring decisions (Gottfredson & Saklofske, 2009; Snyderman & Rothman, 1987; Thorndike, 1997). Tests such as the SAT and GRE are also used for admissions processes into academic programs and financial-aid decision making (Williams, 1996). Although demographics, transcripts, and experience factor into such decisions, there has been much concern about limiting access to higher education and screening out many racial/ethnic minority students (Awad, 2007; Goldstein et al., 2010; Guthrie, 1976; Valencia & Suzuki, 2001).

Some scholars have argued that intelligence tests have been used to reinforce the dominance of mainstream American values (Guthrie, 1976; Ramirez & Gonzalez, 1972; Sanchez, 1932). The APA task force also noted that claims about racial/ethnic differences have been used to justify discrimination in America (Neisser et al., 1996). Western values such as speed, abstraction, analytical and verbal abilities are embedded within intelligence measures such as the WAIS-IV (Sternberg, 2000; Valencia & Suzuki, 2001). Ramirez and Gonzalez (1972) noted that in spite of cultural differences in learning styles, intelligences tests tend to cater to the cognitive styles of European Americans. Several researchers have emphasized the role of language, noting that Mexican Americans have often scored higher when retested in Spanish (Sanchez, 1932; Guthrie, 1976; Olmedo, 1981; Ramirez & Gonzalez, 1972). In addition, many of the items on intelligences tests, both verbal and nonverbal, are culturally loaded (Ramirez & Gonzalez, 1972; Valencia &

Suzuki, 2001). Although nonverbal subtests are often regarded as "culturally reduced," exposure to puzzles, video games, educational toys, and other cultural phenomena can familiarize test takers with items on intelligence tests, which can influence results (Murdoch, 2007; Ramirez & Gonzalez, 1976).

The potential misclassification of racial/ethnic minority students has roused great controversy, peaking in 1979 with the case of *Larry P. vs. Riles*, which resulted in a ban against the intelligence testing of African Americans in California. Scholars have noted that intelligence tests have misclassified African American and Latino American students as educably mentally retarded and that racial/ethnic minority students continue to be overrepresented in special education (Buss, 1996; Power, Hagans-Murillo, & Restori, 2004; Ramirez & Gonzalez, 1972). This decision was appealed and eventually overturned in 1994 after several African American families protested that denying their children needed services was discriminatory (see Buss, 1996 for a thorough discussion). While the situation in California was extreme, the issues remain salient to racial/ethnic minorities and the mental health care professionals responsible for their assessment. Thus, continued research is needed to understand the various factors contributing to racial/ethnic differences on cognitive ability measures and academic achievement.

Summary of Stereotype Threat Findings

Stereotype threat is defined as a sociopsychological threat evoked by an evaluative situation that can confirm a negative stereotype about one's group (Steele, 1997). In a series of four seminal studies, Steele and Aronson (1995) provided the first evidence of the presence of stereotype threat. They theorized that the stereotype of

African Americans as being low in intellectual ability could lead to fears of confirming this negative stereotype, which could in turn interfere with performance. In Study 1, Steele and Aronson administered the Graduate Record Examination (GRE) as a measure of cognitive ability to both African Americans and European Americans under diagnostic and non-diagnostic conditions. In the diagnostic condition meant to arouse stereotype threat, participants were told that the test was a genuine assessment of their verbal abilities and that they would receive feedback on their strengths and weaknesses. However, in the non-diagnostic condition, participants were told that researchers were only interested in psychological variables related to problem-solving and that participants could benefit by being exposed to the types of problems they might encounter on future tests. Participants were asked to give their best effort despite the fact that their abilities were not being assessed. Participants in both conditions were told that they should not expect to answer many items correctly due to the difficulty of the test.

Although the same GRE was used in all conditions, the manner in which the test was presented produced significantly different results. In the diagnostic condition, the mean African American score was a standard deviation below that of the mean European American GRE score, reminiscent of the achievement gap that has often appeared in studies of standardized test performance (Davis & Simmons; 2009; Herrnstein & Murray, 1994; Steele & Aronson, 1995). However, in the non-diagnostic condition, African Americans performed significantly better than their counterparts in the diagnostic condition, and were not significantly different from their European American peers in the non-diagnostic condition. Study 2 replicated the effects of Study 1 and also demonstrated

that African Americans in the diagnostic condition on average worked more slowly and were less accurate than African Americans in the non-diagnostic condition.

In Study 3, results suggested that stereotype activation had taken place in the diagnostic condition. A word association task revealed that race-related stereotypes had been primed, such that African Americans in the diagnostic condition produced more race-related words than did those in the non-diagnostic condition. Furthermore, African Americans in the diagnostic condition were much less likely to indicate their race (25% relative to 100% in the non-diagnostic condition) and less likely to endorse stereotypic activities, such as basketball and listening to rap music. These results provided the first evidence that racial stereotypes are salient for African Americans under stereotype threat conditions. Lastly, Study 4 demonstrated that indicating race prior to taking the GRE disrupted test performance even when the test was not presented as diagnostic. Integrating the results of these studies, Steele and Aronson (1995) argued that contending with negative stereotypes while taking an exam can disrupt African American performance by increasing self-doubt and anxiety.

Several studies have since replicated and extended the theory of stereotype threat to various other subgroups, including Latinos (Armenta, 2010; Gonzales, Blanton, & Williams, 2002; Guyll, Madon, Prieto, & Scherr, 2010; Schmader & Johns, 2003; Van Laar, Levin & Sinclair, 2008), women (Gonzales et al., 2002; Nguyen & Ryan, 2008; Quinn & Spencer, 2001; Schmader & Johns, 2003, Shih, Pittinsky, & Ambady, 1999; Spencer, et al. 1999), athletes (Yopyk & Prentice, 2005), people of lower SES (Harrison, Stevens, Monty & Coakley, 2006; Spencer & Castano, 2007) and even European American men (Aronson, Lustina, Good, & Keough, 1999). Researchers have also

examined stereotype threat in different settings, such as employment seeking (McKay, Doverspike, Bowen-Hilton, & Martin, 2002) and students undergoing the SAT and Advanced Placement (AP) exams (Cullen, Hardison, & Sackett, 2004; Stricker & Ward, 2004). Stereotype threat appears to affect people of all levels of cognitive ability (Steele, 1997). Domain identification is a necessary factor of stereotype threat (Steele & Aronson, 1995; Steele, 1997), which may explain why even the most gifted students can experience reduced performance under stereotype threat conditions. It follows that if a given ability is important to one's self-image, then distress arises when the ability is drawn into question. Most stereotype threat studies have focused on college students, who represent an ostensibly high-achieving group. For example, Good, Aronson, and Harder (2007) found that very qualified and persistent college women enrolled in advanced calculus courses were vulnerable to stereotype threat. Thus, it is clear that stereotypes can and do influence test performance for almost any group.

Individuals often belong to multiple groups, which has important implications for stereotype threat. Whereas some groups (e.g., African Americans, Latino Americans, athletes) have negative stereotypes associated with their intellectual ability, other groups, (e.g., European Americans, Asian Americans, college students) are stereotyped to be intelligent. When individuals fall into both of these categories, their performance can vary according to the salience of relevant stereotypes. Shih et al. (1999) demonstrated that Asian women performed better on a standardized math test when their Asian identity was primed yet performed worse when their gender was made salient. Similarly, Yopyk and Prentice (2005) found that the performance can also be affected as a function of identity prime; student-athletes performed worse when reminded of their athlete status,

but showed no performance decrement when their student identity was primed. Capitalizing on the fact that most people have multiple social identities, Rydell, McConnell and Beilock (2009) found that presenting a positive association in conjunction with a negative one had the effect of reducing stereotype threat. Together these findings demonstrate that group membership and group-based stereotypes can have important implications for performance on achievement and cognitive ability tests.

While multiple group membership can have protective effects, belonging to more than one stigmatized group can exacerbate stereotype threat. Gonzales et al. (2002) demonstrated that having double-minority status compounded the effects of stereotype threat. They found that stereotype threat affected the performance of Latinos more than that of European Americans and that Latinas experienced a greater performance decrement than Latino males. Gender and race/ethnicity had an interactive effect, placing ethnic women more at risk for stereotype threat. However, the effect was asymmetrical, such that the race/ethnicity-based stereotypes sensitized Latinas to gender-based stereotypes, and gender-based stereotypes did not appear to sensitize them to ethnicity-based stereotypes. These findings suggest that when stereotypes related to race/ethnicity are more salient, women of color are more vulnerable to stereotype threat than European American women. Measures of cognitive ability often include subscales of arithmetic, a domain in which women and racial/ethnic minorities are stereotyped to perform more poorly than European American men (Quinn et al., 2001). Research has also consistently reported that men tend to be better at spatial reasoning than women (Martens, Johns, Greenberg, & Schimel, 2006) which could also contribute to stereotype threat during the performance tests on the WAIS-IV. Together, these findings demonstrate the importance

of examining the presence of stereotype threat in testing situations in which intellectual ability is being assessed.

Cultural Identity and Stereotype Threat

Awareness of the long-standing IQ gap between European Americans and certain racial/ethnic minority groups is likely to cause apprehension among stereotype-vulnerable populations. Yet given that members of racial/ethnic minority groups are likely to identify with their native cultures to varying degrees, racial/ethnic identification is an important variable when considering racial/ethnic minority performance on intelligence tests. Considering varying definitions of cultural identity in the literature, a brief discussion of terminology is warranted.

Cultural identification has typically been operationalized in terms of racial identity and ethnic identity, which although sometimes used interchangeably, have different conceptual meanings. Racial identity refers to the degree of identification with one's racial group. In the literature, race and racial identity are most often discussed in terms of African American and European American differences because of the contentious historical background of these groups. Furthermore, some have argued that African American culture is often not as readily identified as that of Latinos or Asians because the African American population originates from within the United States and speaks English (Landrine & Klonoff, 1996). This perspective has been demonstrated in the literature, with far fewer studies examining African American ethnic identity than racial identity.

In contrast to racial identity, ethnic identity is more commonly studied in Latino American and Asian American populations (Kim & Abreu, 2001). Ethnic identity refers to the degree of identification to one's ethnic group, which is based on shared ancestry, history, and cultural characteristics such as values, beliefs, customs, and language (Smedley & Smedley, 2005). Ethnic identity emerges through the processes of enculturation and acculturation. The process of enculturation involves socialization to the culture of origin whereas acculturation is conceptualized as adaptation to new cultures, which may occur through assimilation (losing one's cultural heritage), biculturation (belonging to both the heritage and new cultures), marginalization (belonging to neither culture) (Berry, 2005; Kim & Abreu, 2001). Individuals may also choose not to acculturate, which is sometimes described as separation (Berry, 2005).

Phinney (1996) operationalized ethnic identity based on Marcia's (1966) theory of identity development and focused on three dimensions of ethnic identity: exploration, the degree to which one seeks knowledge about one's heritage, affirmation, which represents an individual's feelings about being a member of his/her ethnic group, and resolution, the degree to which an individual has a clear sense of what being their ethnic group membership means to them. Her measure is known as the Multigroup Ethnic Identity Measure (MEIM), which has been widely used in the literature. Umaña-Taylor, Yazedjian, and Bámaca-Gómez (2004) critiqued the MEIM, disagreeing with the practice of collapsing the three subscales into one total score, instead arguing that Marcia's theory was best represented by three distinct dimensions. In addition, Umaña -Taylor et al. noted that the affirmation items for the MEIM were all positively worded and did not allow for

the possibility that one may have negative feelings about their ethnic group. They addressed these issues with the development of the Ethnic Identity Scale (EIS).

Both the MEIM and EIS were designed to be applicable to all people, following Phinney's (1996) assertion that everyone has an ethnic identity. As a result, ethnic identity is often used for cross-cultural research. In contrast, racial identity measures tend to be designed explicitly for African Americans. Thus, in the current study ethnic identity was examined to facilitate comparisons between African Americans, Latino Americans, and European Americans. In addition, while differences on cognitive ability tests are usually referred to as racial differences, some of the explanations for the achievement gap are cultural in nature. As with African American identity (Cokley, 2006), the achievement gap has often been racialized. However, this language minimizes the role of cultural variables. Thus, exploration of ethnic identity to cognitive test performance is needed. However, few studies specifically examine the role of ethnic identity in stereotype threat; racial identity and acculturation have been more commonly studied in the context of cognitive test performance. Thus, based on the conceptual overlap of racial and ethnic identity with regard to cultural elements, the extant literature of the relationship of racial identity and acculturation to stereotype threat were also considered to inform the current study.

Research has produced mixed findings regarding the roles of racial/ethnic identity to cognitive performance. Some research has found that the effect of stereotype threat is moderated by the continuum of racial identity (Davis, Aronson, & Salinas, 2006; McFarland et al., 2003; Ployhart et al., 2003; Sellers, Chavous, & Cooke, 1998). It follows that if one does not identify with one's racial/ethnic group, a negative stereotype

about the group may be perceived as irrelevant and thus unthreatening. On the other hand, strong racial/ethnic identification may render negative stereotypes more threatening. Armenta (2010) recently found that highly ethnically-identified Latino Americans performed worse whereas the performance of lower identified Latinos was not affected by stereotype threat. Similarly, Ployhart et al. (2003) found that African Americans who endorsed high racial identity performed significantly worse under stereotype threat conditions compared to African Americans who endorsed lower racial identity. In addition, there is some evidence to suggest that more acculturated members of ethnic minority groups tend to perform better on cognitive ability tests. Olmedo (1981) found that more acculturated Latino Americans score higher on tests of cognitive ability, suggesting that such scores may reflect linguistic acculturation rather than intelligence.

In a different vein of thinking, it stands to reason that racial/ethnic minority persons who have higher racial identity could be protected against stereotype threat. Those who identify strongly with their racial/ethnic group may derive support, strength, and a will to dispel negative stereotypes about their group (Davis et al., 2006; Sellers et al., 1998).

In contrast to studies that have suggested that racial/ethnic identity can have a powerful influence on performance, other researchers have failed to find an association between racial/ethnic identity and academic achievement or test performance (Awad, 2007; Davis & Simmons, 2009; Smith & Hopkins, 2004). Smith and Hopkins (2004) found no differences between high and low acculturation on academic performance. In addition, Awad (2007) found that racial identity attitudes did not predict GPA or GRE scores. It has been argued that more proximal variables such as academic self-concept are more relevant to academic achievement, suggesting that racial/ethnic identity may be too

distal to have a direct effect on test performance (Awad, 2007; Sellers et al., 1998). Furthermore, while Manly, Byrd, Touradji, and Stern (2004) initially found that more acculturated elderly African Americans outperformed more traditional African Americans on neuropsychological measures, the effects of acculturation were largely diminished once demographic characteristics, such as age, sex, and education, were taken into consideration.

Methodological variations may contribute to the conflicting findings regarding the relationship of racial/ethnic identity and test performance. Awad (2007) noted that some of the discrepant findings may be related to the use of different assessment measures of racial/ethnic identity. For example, some studies have used an older racial identity measure, the Racial Identity Attitude Scale (RIAS; Davis et al. 2006; McFarland et al., 2003), while others have used the more recent Cross Racial Identity Scale (CRIS; Awad, 2007), which has better psychometric properties, while still others have used the Multidimensional Inventory of Black Identity (MIBI; Sellers et al., 1998) or the MEIM (Armenta, 2010).

Timing of demographic measures is another important methodological consideration in the context of examining stereotype threat. Several studies have induced stereotype threat by asking participants to complete racial/ethnic identity measures prior to an exam (Brown & Day, 2006; Davis et al., 2006; McFarland et al., 2003; Steele & Aronson, 1995). Interestingly, McFarland et al. (2003) noted that the timing of racial identity measures can affect the interpretation of results, which may also contribute to mixed findings in the literature. In one condition of the McFarland study, participants completed a racial identity measure prior to taking a cognitive ability exam, and in the

other, they completed the racial identity measure after the exam. Overall, participants who endorsed higher racial identity performed worse on the test. However, a comparison of pre-test and post-test racial identity revealed that test-takers that indicated lower racial identity post-test performed better, suggesting that that those who disidentified from their racial groups, and thus the negative stereotype, performed better. If only pre-test scores were examined, one would conclude that African Americans with high racial identity were buffered from stereotype threat. Yet if only post-test scores were considered it would appear that higher racial identity is associated with poorer test scores. Thus, it is important to consider the timing of racial/ethnic measures when interpreting results of stereotype threat studies.

Thus, it appears that the act of taking a cognitive ability exam can influence the expression of one's racial identity. Other researchers have noted a similar trend of racial/ethnic disidentification in order to distance themselves from negative stereotypes. Steele and Aronson (1995; Study 3) observed this phenomenon, noting that African Americans participants under stereotype threat conditions were less likely to admit enjoying stereotypical cultural activities. Cohen and Garcia (2005) also found that participants vicariously exposed to stereotype threat avoided endorsing cultural stereotypes, inhibited stereotype-related thoughts, and physically distanced themselves from their threatened peers. These findings suggest that racial/ ethnic minority test-takers manage negative feelings during exams by disidentifying from their ethnic group, suggesting that they are attempting to evade stereotype threat.

Another important methodological consideration for the test performance of racial/ethnic minority persons is the race/ethnicity of the test administrator. Some

racial/ethnic minority students may be socialized to be wary of European Americans. Terrell, Terrell, and Taylor (1981) have labeled this concept as cultural mistrust, which is thought to serve a protective function against racism and discrimination. Furthermore, some researchers have found that racial/ethnic match between administrator and examinee reduces stereotype threat and that racial/ethnic mismatch impairs performance (Marx & Goff, 2005; Terrell et al., 1981). Marx and Goff (2005) found that African Americans tested by European Americans underperformed relative to African Americans tested by African Americans. Experimenter race/ethnicity had no effect on European American test takers. Marx and Goff argued that the presence of an African American examiner changes the meaning of the test and reduces the likelihood of being stereotyped threat. In addition, African American participants tested by European American examiners were more likely to endorse the stereotype that African Americans were not as intellectually capable as European Americans than African Americans tested by African American examiners. Although agreement with the stereotype is not a necessary condition for stereotype threat (Steele & Aronson, 1995), believing negative stereotypes heightens the level of threat. Marx and Goff found that African Americans tested by European Americans reported higher levels of stereotype threat and that belief in negative stereotype mediated the relationship between experimenter race and performance. Although these findings suggest that racial/ethnic mismatch can impair performance, other researchers have found that experimenter race/ethnicity does not affect intelligence test scores, (Sappington & Grizzard, 1975; Wout, Shih, Jackson, & Sellers, 2009).

Wout et al. (2009) offer insight into these conflicting findings by focusing on the perception of stigmatized participants. In a series of five studies, they demonstrated that

group membership of the evaluator was only relevant when test takers had reason to believe that they might be stereotyped. When participants were tested by European Americans who endorsed negative stereotypes about their group, scores were negatively affected by stereotype threat. However, when European American test administrators shared that they did not endorse the threatening stereotype, stereotype threat did not take place. The endorsement of stereotypes by African American test administrators mirrored these effects. Thus, perceptions of the examiner's endorsement of negative stereotypes predicted the effects of stereotype threat. Wout et al. (2009) suggested that in the absence of individuating information about the evaluator, test subjects may resort to group membership stereotypes, which is that White examiners would expect African Americans to be less intelligent.

Situational factors are also important to consider with regard to the relationship of racial/ethnic identity and stereotype threat. For example, the level of stereotype threat present in a situation has been found to affect the influence of racial identity. Davis et al. (2006) examined the relationship of different racial identity attitudes to test performance. They found that racial identity attitudes were influential at low and moderate levels of stereotype threat. However, at high levels of stereotype threat, racial identity attitudes did not predict performance; all participants performed more poorly in the high threat condition than in the moderate and low threat conditions. The authors argued that, as is commonly found in social psychology, personality factors such as racial/ethnic identity tend to have the most influence when situational demands are ambiguous or low. Yet when situational demands are strong, personality factors can be overridden.

Exploring test taker perceptions of the WAIS-IV could provide insight into the situational demands of the WAIS-IV. On the one hand, race and ethnicity are not typically discussed during WAIS-IV assessment; thus, these stereotypes may not be particularly salient given that protocol measures explicitly avoid the use of the word intelligence to prevent anxiety. On the other hand, when racial/ethnic minority persons enter a testing environment in which they are clearly being assessed, the evaluative situation itself may evoke stereotype threat (Steele, 1997). Racial/ethnic identity may also contribute to perceptions of the testing environment. More racially/ethnically identified participants may be more likely to anticipate being stereotyped, rendering them more vulnerable to stereotype threat, whereas less identified participants may not be attuned to the potential threat. The current study sought to provide insight regarding the role of ethnic identity and test performance.

Stereotype Threat and Anxiety

In order to better understand how best to reduce stereotype threat, it is necessary to consider the mechanisms through which it functions. Several potential mediators of stereotype threat have been examined, including negative thoughts (Cadinu, Maass, Rosabianca, & Kiesner, 2005), disruption of problem-solving strategy formation (Spencer Steele, & Quinn 1999), decreasing accuracy on test performance (Brown & Day, 2006; Marx & Goff, 2005; Shih et al., 1999; Steele & Aronson, 1995), and procedural memory (Beilock, Jellison, Rydell, McConnell & Carr, 2006). While many potential intervening variables in the process of stereotype threat have been studied, results have been mixed

(David & Simmons, 2009). The current study focused on the potentially mediating role of anxiety.

Anxiety is a plausible mediator for stereotype threat in performance on the WAIS-IV because having one's intelligence evaluated can evoke fear, self-doubt, and distress. Since the introduction of stereotype threat, Steele and Aronson (1995) have suggested that stereotype threat operates by increasing anxiety. They posited that anxiety acts as a distracter from the task at hand, shifting attentional resources away from problem-solving. However, the literature has produced mixed results regarding the relationship of anxiety and stereotype threat (Ryan & Ryan, 2005; Steele & Aronson, 1995; Steele & Aronson, 2004), with some studies finding a positive relationship between stereotype threat and anxiety (Harrison et al., 2006; Osborne, 2001) and others finding inconclusive results (Spencer et al., 1999) or no association (Mayer & Hanges, 2003). Test anxiety in particular is likely to contribute to stereotype threat due to the focus on evaluation, but findings are also mixed with regard to this relationship (Aronson et al., 1998; Spencer et al., 1999; Steele & Aronson, 1995).

These findings suggest that the relationship of test anxiety to stereotype threat may be complex, perhaps functioning as an intervening variable. Ryan and Ryan (2005) proposed a model that integrated stereotype threat and achievement goal theory, arguing that stereotype threat evokes a performance-avoidant goal, in which the aim is to avoid negative judgment of ability and feelings of incompetence. Subsequently, worry and test anxiety are triggered, which threaten self-efficacy and diminish cognitive capacity. Thus, this rationale suggests that test anxiety mediates the relationship between stereotype threat and test performance.

Sawyer and Hollis-Sawyer (2005) argued that test anxiety has a place as a mechanism by which stereotype threat occurs, either directly or indirectly through both person and situation characteristics. Supporting this claim, Osborne (2001) found that post-test anxiety partially mediated racial/ethnic differences on performance on achievement tests. While Osborne argued that these findings were consistent with the predictions of stereotype threat theory, there was no manipulation or assessment of stereotype threat in this study. However, in a later study (Osborne, 2007), he found that under high stereotype threat conditions, women exhibited increased physiological arousal, such as heart rate, skin conductance, and diastolic blood pressure. Jordan and Lovett (2007) also noted that physiological anxiety has been clearly linked to stereotype threat. However, more research is needed to help resolve conflicting findings in the literature. Thus, the current study contributed to the existing literature by investigating two specific types of anxiety, state and test anxiety, as possible mediators of stereotype threat in the WAIS-IV.

Reducing Stereotype Threat

In addition to literature focused on understanding the mechanisms behind stereotype threat and identifying variables that can affect its presentation, a substantial body of literature has focused on different strategies to reduce stereotype threat for vulnerable populations (Nguyen & Ryan, 2008). Stereotype threat theory predicts that reducing stereotype threat should improve the performance of stereotyped individuals (Steele, 1997). Therefore, reducing stereotype threat on intelligence exams could increase their validity by reducing potential bias, decreasing anxiety, and removing the burden of

negative stereotypes. Previous studies have reduced threat by priming positive stereotypes (Rydell et al. 2009; Shih et al. 1999; Yopyk & Prentice, 2005), informing participants about stereotype threat (Johns, Schmader, & Martens, 2005; Williams, 2005), emphasizing overlapping characteristics between stereotyped and non-stereotyped groups (Rosenthal & Crisp, 2006; Rosenthal, Crisp, & Suen, 2007), and self- affirmations (Cohen, Garcia, Apfel, & Master, 2006; Martens et al., 2006). The current study focuses on the stereotype threat reduction strategy of emphasizing the malleability of intelligence, which has be found to be effective by several researchers (Aronson et al., 2002; Aronson & Steele, 2005; Blackwell, Trzesniewski, & Dweck, 2007; Good, Aronson, & Inzlicht, 2003).

STEREOTYPE THREAT REDUCTION THROUGH MALLEABILITY

Individuals' beliefs about the nature of intelligence have important implications for the effects of stereotype threat. For individuals who believe that intelligence is endowed at birth and cannot be substantially changed, negative stereotypes are more potentially threatening. On the other hand, believing that intelligence is expandable or malleable could encourage stigmatized individuals that despite negative stereotypes, they have the potential to improve and succeed. The belief that human traits, such as intelligence, are malleable is known as an incremental theory, whereas believing that intelligence is a finite quality endowed at birth is referred to as an entity theory (Dweck & Leggett, 1988).

Dweck and Leggett (1988) proposed a motivational model of achievement, in which self-theories of intelligence affect responses to academic challenges by influencing motivation to improve, which can have long-term implications for performance. The

model proposes that after performing poorly, incremental theorists are likely to be motivated to improve because they perceive their goals as attainable. In contrast, entity theorists may believe such improvement is not feasible and thus may feel hopeless. Students who hold incremental beliefs about intelligence have been found to earn higher grades and value learning and mastery, rather than simply performing well (Blackwell et al., 2007; Dweck & Leggett, 1988; Dweck, Chiu, & Hong, 1995). In contrast, entity theorists may be more at risk for negative outcomes, such as feeling helpless in the face of adversity (Dweck & Leggett, 1988).

Although implicit theories of intelligence are informed by socialization and life experience, implicit theories can be temporarily and experimentally manipulated (Aronson et al., 2002; Burns & Isbell, 2007; Nussbaum & Dweck, 2008). Burns and Isbell (2007) argued that implicit theory priming can induce participants to behave similarly to incremental theorists, regardless of their pre-existing theory. Furthermore, Nussbaum and Dweck (2008) demonstrated that when experiencing a threat to self-esteem after poor performance, participants primed with incremental theories were more driven to improve their performance than those primed with entity theories. Due to the fact that implicit theories are amenable to change, they present an opportunity to reduce stereotype threat.

Drawing on the concept of implicit theories, Aronson et al. (2002) theorized that emphasizing malleability can result in improved academic achievement. They reasoned that stereotype threat could lead entity theorists and stigmatized individuals to be overly focused on performance, which could lead to increased pressure to do well and anxiety. Aronson et al. aimed to prevent these negative sequelae by encouraging stigmatized

individuals to adopt incremental goals (e.g. learning) rather than performance goals when encountered with stereotype threat. To this end, they conducted an intervention in which college students mentored fictional middle school students as pen pals in three sessions. Those in the control condition were instructed to write generally encouraging statements to their pen pals. However, those in the malleability condition received information about the malleability of intelligence and were encouraged to pass these messages along to their mentees. Thus, the writing sessions served as an intervention to indoctrinate the concept of malleability. After three sessions of writing, students in the malleability condition reported more enjoyment of school, more academic engagement, and higher grades compared to students in the control condition. They found the intervention helpful for both African American and European American students, although more effective for African Americans.

Other researchers have also carried out successful malleability interventions in academic settings. Blackwell et al. (2007) conducted a successful intervention that instructed a diverse group of junior high school students about the incremental theory of intelligence. The students who received this message outperformed those in the control group, whose grades took a downward turn. Good et al. (2003) also demonstrated that a malleability intervention resulted in the improved performance of junior high girls and Latino adolescents. Both girls and boys benefitted from the malleability condition, but the increase was higher for girls. Furthermore, math and reading standardized achievement test scores increased for those in the malleability condition. In addition, Aronson (1998) reviewed a series of studies conducted by himself and his colleagues that aimed to reduce stereotype threat with a focus on learning processes and malleability. Aronson argued

that if participants have experiences of success and improvement, stereotype threat could be reduced. These studies were promising, finding that after malleability interventions, participants had improved performance on the GRE, less test performance anxiety, and identified more strongly with academics. However, many of these were unpublished manuscripts (Aronson & Fried, 1997; Aronson & Salinas, 1997; Aronson, Tichy, & Croteau, 1997) and thus must be considered tentatively. It is necessary for malleability research to reach a wider audience in order for greater effects to be achieved. Following the reasoning of Aronson et al., if racial/ethnic minority students adopt a performance goal mentality in response to stereotype threat, their performance on intelligence tests could be affected by higher endorsement of stereotype. In the context of intelligence testing, priming incremental theories could serve to increase motivation to improve performance.

As noted, the majority of malleability interventions have been conducted in academic contexts. This author identified only one study that discussed the role of malleability in test performance on an intelligence test. Using culture-fair matrices as a measure of intelligence, Sawyer and Hollis-Sawyer (2005) found that belief in the malleability of intelligence was inversely related to stereotype threat. Malleability was also significantly associated with test anxiety. However, malleability was a correlational variable and was not manipulated. Experimental studies are needed to understand whether malleability interventions may also be successful in clinical settings. The current study aimed to extend and strengthen these findings by more clearly delineating the relationships of test anxiety, stereotype threat and performance on intelligence tests.

Many researchers acknowledge that the stereotype threat reduction strategies thus far proposed may be impractical, in that they are not likely to be employed in the real world (McKay et al., 2002; Nguyen & Ryan, 2008). Telling participants that tests are not meant to be diagnostic or gender fair could be misleading and unethical. However, given that a substantial body of literature exists for the effectiveness of the malleability of intelligence strategy, this approach does not present any ethical dilemmas and could be implemented in testing situations.

Criticism of Stereotype Threat

Despite wide replication of Steele and Aronson (1995)'s findings, some researchers remain skeptical of its real-world application (Cullen et al., 2004; Sackett, Schmitt, Ellington, & Kabin, 2001; Stricker & Ward, 2004; Sticker & Ward, 2008). Sackett et al. (2004) argued that the original findings have been widely misinterpreted and misconstrued in both popular and scientific writings. They stated that although Steele and Aronson did not claim to completely explain racial/ethnic test score differences, stereotype threat is widely claimed to do so in the literature. They criticized Steele and Aronson because they controlled for prior SAT scores, and argued that no-threat conditions simply replicated pre-existing differences. In other words, they argued that stereotype threat merely exacerbated existing test score differences.

Steele and Aronson (2004) criticized Sackett et al. (2004)'s interpretation as narrow, and encouraged readers to consider the broader stereotype threat literature as a whole, which has consistently replicated stereotype threat, at times with covarying SAT scores, but many times without doing so (Cohen & Sherman, 2005). Steele and Aronson

(2004) disagreed with the fact that stereotype threat has been widely misunderstood, acknowledging the examples that Sackett et al. (2004) pointed out, but arguing that a far greater number of studies correctly referenced stereotype threat as contributing to, but not fully explaining the test score discrepancy (see also Cohen & Sherman, 2005). Steele and Aronson (2004) argued that stereotype threat conditions reflect real-world testing. They noted that the traditional test gap between African Americans and European Americans is replicated under stereotype threat conditions, suggesting that the laboratory results mirrored what occurs in actual testing situations. Furthermore, Cohen and Sherman (2005) argued that studies conducted in real-world settings, such as classrooms, have led to true and lasting effects, giving further credence to the benefits of stereotype threat reduction.

Thus, it is clear that the ecological validity of stereotype threat theory has been challenged. The few studies of stereotype threat that have been conducted in actual high-stakes testing situations have produced mixed findings (Sackett et al., 2001; Stricker & Ward, 2008). Stricker and Ward (2008) remarked that of the six studies of stereotype threat in actual high-stakes testing situations of which they were aware, five failed to replicate the positive laboratory findings. They maintained the findings of Stricker and Ward (2004) which concluded that changing the placement of demographic information before or after test questions had no effect on the performance of women and ethnic minorities taking Advanced Placement exams. Other researchers (Huguet & Régner, 2007; Wicherts et al., 2005) have reported similar results. However, Danaher and Crandall (2008) argued that Stricker and Ward (2004) used overly conservative criteria to evaluate the practical significance of the procedural modification. Danaher and Crandall

reanalyzed the data of Sticker and Ward (2004) concluding instead that there was a significant benefit for women that answered demographic questions after the AP exam, with greater rates of passing the exam, fewer gender differences in test scores, translating to an increase in women receiving AP credit. They acknowledged that the protocol modification did not appear beneficial for African Americans, stating that perhaps the subtlety of the manipulation was too weak to influence the salience of such prevalent stereotypes in the testing environment.

Similarly, Jordan and Lovett (2007) argued that the Stricker and Ward (2004) study was a poor test of stereotype threat's relevance to the real world, given that stereotypes associated with race, gender, and ability are so prevalent that completing demographic measures may not have increased stereotype activation. These findings demonstrate the large empirical gap between the many laboratory replications of stereotype threat and the lack of evidence of its existence in real-world settings. The current study sought to close this gap by evaluating whether stereotype threat existed under normal testing conditions and whether this threat could be reduced by emphasizing the malleability of intelligence.

Limitations of Stereotype Threat Literature

Several limitations exist in the studies of stereotype threat in the current literature. Most research has been conducted on women's performance on math tests. Despite the fact that the original study of stereotype threat focused on African Americans, there are far fewer studies on African Americans than women (Sackett, Hardison, & Cullen, 2004b). Research focused on Latino Americans has been even scarcer (Armenta, 2010;

Gonzales et al., 2002). Furthermore, Nguyen and Ryan (2008) reported that stereotype threat removal strategies were less effective for racial/ethnic minorities than for women. Therefore, more research is needed to better understand how to successfully reduce stereotype threat in clinical settings for people of color. Given the stereotype of low intelligence among African Americans and Latino Americans and the relative lack of research in this area, the current study focused on these racial/ethnic minority groups.

Some may argue that there is a great degree of overlap between intelligence and academic achievement tests, but there are several notable differences. Thus, another limitation of the stereotype threat literature is the preponderance of research focused on academic settings. Steele and Aronson (1995) began the examination of stereotype threat in the context of the GRE and other standardized testing instruments. Accordingly, the vast majority of stereotype threat literature has focused on academic achievement. However, relatively little research exists on stereotype threat in psychological assessment contexts or commonly used measures of intellectual functioning (Jordan & Lovett, 2007). Some researchers have used Raven's Advanced Progressive Matrices to demonstrate the effects of stereotype threat (Brown & Day, 2006; Croizet et al. 2004, Mayer & Hanges, 2003; McKay et al., 2002). However, a review of the literature by the author did not reveal any published studies using the WAIS-III or WAIS-IV, although the WAIS exams are the most widely used tests for academic placement, intelligence testing, and for disability service qualifications in the field of psychology.

Although both the SAT and WAIS share their origins in the Army Alpha, there are important differences in their administration procedures and applications, which may affect the influence of stereotype threat. For example, standardized tests often

administered in large groups for the sake of efficiency whereas intelligence tests are typically administered individually. Item presentation also varies across the two types of evaluations. The items of the SAT are all multiple choice, with the exception of the writing section. In contrast, several of the WAIS-IV items are open-ended, often requiring examinees to generate verbal responses. An oral examination by a test administrator may be more intimidating than a written exam that does not require such interaction.

The SAT and WAIS-IV are also used for related although distinct purposes. While intelligence tests aim to asses a wide variety of cognitive abilities, including both verbal and nonverbal components, the SAT is limited to literate participants and assesses verbal, quantitative, and writing abilities. Intelligence testing is typically reserved for the extremes of the ability continuum, used to assess learning disabilities, mental retardation, or giftedness. Another consideration is the accessibility of such tests. For the SAT, there are preparation classes and college-bound curricula that aid in success on the SAT (Ceci, 1996). In contrast, the test security of intelligence tests is highly valued, in an effort to keep measurements "pure." Due to the many differences between standardized tests and intelligence measures such as the WAIS-IV, more research is necessary with tests specifically designed to assess intelligence to understand how stereotype threat may operate in these contexts.

Another limitation of the stereotype threat literature is that nearly all studies manipulate threat by induction through priming stereotypes (Aronson et al. 1999; Steele & Aronson, 1995; Spencer et al. 1999), priming particular abilities and skills (Shih et al., 1999; Yopyk & Prentice, 2005), and manipulating diagnosticity (Croizet et al., 2004;

Gonzales et al., 2002; McKown & Weinstein, 2003; Spencer & Castano, 2007; Steele & Aronson, 1995). However, as some researchers (Davis et al., 2006; Sackett et al., 2001) have noted, this methodology lacks ecological validity because purposeful induction of stereotype threat is not conducted in real testing environments. However, Brown and Day (2006) provided evidence that a level of threat exists even in standard administration procedures, and that while African Americans underperformed relative to European Americans in a standard threat condition, in which racial/ethnic primes were not used. Other researchers have also demonstrated that even when stereotypes were not stated explicitly or manipulated, women's performance on math tests was still affected (Huguet & Régner, 2007; Quinn & Spencer, 2001; Spencer et al., 1999). It has been argued that standard administration protocols can naturalistically evoke substantial stereotype threat. These researchers demonstrated that in the absence of primed stereotype threat, reducing stereotype threat improved performance. Together these studies provide evidence for Steele's statement that stereotype threat lingers as an omnipresent "a threat in the air." Huguet and Régner (2007) stated that while the stereotype threat literature suggests that stereotype threat may be the "default mindset" of stereotyped targets, more empirical research is needed to support this position.

The Current Study

The current study sought to replicate and extend the previous findings of stereotype threat and stereotype threat reduction by emphasizing the malleability of intelligence. Several of the limitations of the current literature were addressed. One asset of the current study was that standard WAIS-IV administration protocol was utilized in

order to understand the level of stereotype threat that occurs naturalistically. Furthermore, this study contributed to the limited body of intelligence assessment using the WAIS-IV. Additionally, it sought to clarify the roles of ethnic identity and anxiety to stereotype threat to WAIS-IV performance in a cross-cultural context. The hypotheses were as follows:

(1) African American and Latino American college students were expected to experience more stereotype threat in the control condition than in the malleability condition, evidenced by higher perceived stereotype threat, higher stereotype vulnerability, and higher salience of race/ethnicity.

(2) African American and Latino American college students were expected to perform better on the WAIS-IV in the malleability condition than in the control condition.

    a. African American and Latina American women were expected to have the lowest scores on the WAIS-IV in the control condition.

(3) Ethnic identity was expected to moderate the relationship between perceived stereotype threat and WAIS-IV performance, such that more ethnically identified participants would have higher levels of stereotype threat and score lower in the control condition than in the malleability condition. Participants who reported less ethnic identification were not expected to differ between conditions.

(4) Test anxiety and state anxiety were expected to mediate the relationship between perceived stereotype threat and WAIS-IV performance in the control condition. (Note: It is recognized that since all variables were measured at post-test,

temporal precedence could not be established). This relationship was not expected in the malleability condition due to the reduced level of stereotype threat.

METHOD

Participants

The sample consisted of 139 undergraduate college students who met the following criteria: (a) considered themselves to be of African American, European American, or Latino American origin/descent and (b) were 18 years of age or older at the time of the study. International students were excluded due to the fact that international students may be subjected to different stereotypes than American students. Participants were recruited from the Introductory Psychology subject pool and an African American psychology class. Data collection began in October 2009 and human subject involvement in this project ended in July 2010.

Twenty-one percent of the sample identified as African American (17 females; 12 males), 41% percent of the sample identified as Latino American (34 females; 23 males) and 38% was European American (20 females; 32 males). One female identified as multi-racial (African American, Native American, and Latino) and was not included in analyses. Thus, the final sample consisted of 138 students. The sample was 52% female. See Table 1 (Appendix O) for more sociodemographic characteristics of the sample.

Experimental Design and Procedure

This study employed a 2(condition) x 3(race/ethnicity). There were two experimental conditions that varied in the amount of stereotype threat: control and malleability. Participants were randomly assigned to one of the two conditions. The racial/ethnic groups were European American, African American, and Latino American.

Due to the fact that previous research has found that the presence of an African American experimenter may reduce stereotype threat (Marx & Goff, 2005; Terrell et al. 1981), WAIS-IV subtests were administered by undergraduate, female, European American research assistants to reduce confounds and examine the effect of the malleability intervention alone.

Participants reported to a designated waiting room, where they were met by a research assistant. Participants signed a consent form at the beginning of the study. Due to the fact that knowing the WAIS-IV is an intelligence test could have affected the results of the study, this information was withheld until participants completed the WAIS-IV, interview, and questionnaires. Furthermore, the WAIS-IV manual encourages test administrators not to mention the word intelligence, as it can cause undue anxiety. Thus, to conceal the true aim of the study, the consent form stated that the purpose of the study was to develop effective test administrators and referred to the WAIS-IV as a series of exercises.

After the participant gave written consent, they were given the State-Trait Anxiety Inventory-state (STAI-state) and Test Anxiety Inventory (TAI) to assess pre-test anxiety levels. Next, experimenters asked participants to read a testing orientation page while testing materials for the WAIS-IV were set up. In the malleability condition, participants were given a malleability prime that described intelligence as a malleable quality that can improve during testing (See Appendix A). In the control condition, participants received a parallel testing orientation to control for time and mental processing that occurs while reading. The control orientation was intended to be redundant with the consent form and to not contain any new information in order to preserve the naturalistic quality of the

study (See Appendix A). There were three manipulation checks to test the effectiveness of the malleability prime, including a questionnaire that assessed to what extent participants endorsed incremental or entity theories of intelligence and two interview open-ended questions that asked participants about their implicit theories of intelligence.

Then, experimenters administered four subtests of the WAIS-IV. In both conditions, experimenters followed the protocol exactly described in the WAIS-IV manual (Wechsler, 2008). However, experimenters gave different verbal testing orientations based on condition. Thus, the written testing orientation was reinforced by the verbal testing orientation. The standard WAIS testing orientation was as follows (brackets indicate the placement of the experimental manipulation in the malleability condition):

I'll be asking you to do a number of things today. Some of the things may be really easy for you, but some may be hard. Most people do not answer every question correctly or finish every item, but please try your best. [Experimental Manipulation) Do you have any questions?

In the control condition, experimenters did not alter the standard WAIS-IV testing orientation and read it to the participants verbatim. In the malleability condition, experimenters read the standard WAIS-IV test orientation and then emphasized the malleability of intelligence by saying:

It is possible to improve your performance as you get accustomed to the test. People have such a large capacity for learning. When you feel challenged, your mind can adapt and develop skills that can help with difficult items.

After the WAIS-IV subtests were administered, the protocol for both conditions was identical. Participants were given the Stereotype Activation Measure immediately after completing the WAIS-IV subtests to assess the presence of stereotype threat (Appendix B). Experimenters conducted interviews to gain a richer understanding of the students' subjective experiences of stereotype threat while participating in the study and more broadly in their academic lives at the University of Texas. The interview questions are listed in Appendix K.

Next, participants completed the questionnaire packet of measures. Finally, experimenters debriefed participants and disclosed the true purpose of the study using the Debriefing Form (Appendix L). Participants then indicated if they wished to remain in the study or withdraw. A few additional interview questions pertaining directly to stereotype threat and intelligence were asked, given that asking these questions prior to the debriefing could have alerted participants to the true purpose of the study (See Appendix M). All participants were given a Resource List (Appendix N) to inform them of resources on campus that they could contact if they experienced any emotional distress.

Measures

Demographics Questionnaire. A general questionnaire was used to gather demographic information such as race/ethnicity, gender, year in college, socioeconomic status, US generational level, college generational level, major, GPA, SAT and ACT scores (See Appendix C).

Wechsler Adult Intelligence Scale-IV (WAIS-IV; Wechsler, 2008). The WAIS-IV is an individually administered clinical instrument designed to assess cognitive ability. The WAIS-IV is a revision of its predecessor, the WAIS-III. It retained 12 subtests of the WAIS-III: Block Design, Similarities, Digit Span, Matrix Reasoning, Vocabulary, Arithmetic, Symbol Search, Information, Coding, Letter-Number Sequencing, Comprehension, and Picture Completion. Three subtests were developed especially for the WAIS-IV: Visual Puzzles, Figure Weights, and Cancellation. General intellectual ability is measured by the Full Scale IQ, which is derived from four subscales that assess specific areas of cognitive functioning: Verbal Comprehension Index, Perceptual Reasoning Index, Working Memory Index, and Processing Speed Index. The WAIS-IV subtests have demonstrated good internal consistency across ages 16-90, ranging from .71 to. 90 (see Table 4.1 in the WAIS-IV Technical and Interpretative Manual for more detailed reliability coefficients by age and subtest; Wechsler, 2008). Interscorer agreement has ranged from .91 to .99. A great deal evidence, based on qualitative and qualitative inquiry of participants' responses, confirmatory factor analysis, and intercorrelations, shows strong support of the validity of the WAIS-IV (Wechsler, 2008).

In the present study, the primary purpose of the WAIS-IV was to assess the level of stereotype threat in a naturalistic testing environment, and not to assess intelligence. Thus, the dependent variables were raw scores of the WAIS-IV subtests rather than IQ scores. For the sake of time, only four of the subtests were administered: Block Design, Digit Span, Matrix Reasoning, and Arithmetic. Thus, two of the tests were in the verbal domain and two were in the performance domain. The Arithmetic subtest was used

because stereotype threat effects have often been shown in tests of mathematical ability. The Digit Span subtest was used because working memory function is hypothesized to be a mediator of stereotype threat. Performance feedback was not disclosed to participants.

Implicit Theory of Intelligence Measure (ITI; Dweck et al., 1995). The ITI is a three-item measure that uses a 6-point Likert scale, ranging from Strongly Agree to Strongly Disagree, intended to assess incremental and entity theories of intelligence. The ITI is scored by averaging responses to the three items. Higher scores indicate higher endorsement of incremental or malleable views of intelligence. High internal reliability has been reported to be .94 - .98 and construct validity has been demonstrated (Dweck et al., 1995). Discriminant validity has been demonstrated and implicit theories are not highly correlated with cognitive ability tests or confidence in intellectual ability (Dweck et al., 1995). In the current study, the ITI served as a manipulation check for the malleability intervention.

Perceived Stereotype Threat (PST; Ployhart et al., 2003). The PST is an eight-item questionnaire that assesses the level of stereotype threat that an individual perceives. It uses a 5-point Likert scale, ranging from Strongly Disagree to Strongly Agree. Responses to PST items are averaged. Higher scores indicate higher perceptions of stereotype threat. The PST can yield a total score, or two subscale scores. The test-specific threat subscale focuses on feelings of stereotype threat related directly to tests. In contrast, the generalized threat scale includes items that measure perceptions of negative attitudes regarding intelligence but do not refer to tests specifically. The PST has been subjected to confirmatory factor analysis, and a two-factor model emerged (Ployhart et al., 2003). In the current study, the total score was used for the sake of simplicity.

Stereotype Activation Measure (SAM; Steele & Aronson, 1995). The SAM is a word completion task that consists of a series of word fragments that may be completed in various ways (e.g. _ _ _ C K). Steele and Aronson (1995) found that under stereotype threat conditions, African Americans produced more race-related completions than under control conditions (e. g. black vs. stack). In the current study, the stereotype activation task was used to measure the presence of stereotype threat.

Race-Based Rejection Sensitivity Scale (RS-Race; Mendoza-Denton, Downey, Purdie, Davis, & Pietrzak, 2002). The RS-Race measures participants' concerns and expectations about being rejected because of their race. It consists of 12 items that describe interpersonal scenarios that could result in race-based rejection. Each item has two sections: the first part asks participants to rate how distressing experiencing the situation would be, and the second part asks about the likelihood of such race-based rejection occurring. Scoring involves calculating the cross-product of the two subscales and averaging across the 12 situations. The RS-Race has been shown to have strong internal reliability, $\alpha = .90$ (Mendoza-Denton et al., 2002). The RS-Race has also demonstrated construct validity in that it correlated with frequency of negative racial interactions, and African Americans have reported higher race-based rejection sensitivity than European Americans (Mendoza-Denton et al., 2002). The RS-Race has also been used as a measure of stereotype vulnerability (Aronson & Inzlicht, 2004).

Ethnic Identity Scale (EIS; Umaña-Taylor et al., 2004). The EIS is a 17-item questionnaire designed to assess ethnic identity. The EIS was derived from Marcia's (1966) typology of identity achievement. It was validated on a large multicultural sample of six ethnic groups (White, Latino, Asian, Black, multiethnic, and other). The EIS

consists of three subscales: (1) exploration, the degree to which individuals have explored their ethnicity, (2) resolution, the degree to which individuals have resolved the meaning of their ethnic identity, and (3) affirmation, the degree to which individuals experience their ethnicity as positive or negative. The EIS is scored by averaging responses to items and examining each subscale separately. Higher scores indicate greater exploration, resolution, and affirmation. The EIS has been found to have good reliability: .89 - .91 for exploration, .84 - .86 for resolution, and .89 - .92 for affirmation (Umaña-Taylor et al., 2004). The EIS has also been shown to have construct validity, demonstrated by the intercorrelations between its subscales and the correlations with familial ethnic socialization and self-esteem measures (Umaña-Taylor et al., 2004). For the current study, ethnic identity was chosen instead of racial identity or acculturation, because ethnic identity facilitates cross-cultural comparisons in a multiethnic sample.

Test Anxiety Inventory (TAI; Spielberger, 1980). The TAI is a 20-item questionnaire that assesses symptoms of test anxiety using a 4-point Likert scale ranging from "1-Never" to "4-Almost Always." The TAI consists of two subscales: TAI-W assesses worry, which is the cognitive component of anxiety, whereas TAI-E assesses emotionality, which is the physiological component of anxiety. High scores indicate higher degrees of anxiety. Both the overall TAI and its subscales have been found to have good reliability, with alpha coefficients ranging from .92 to .96 for the total TAI, .83 to .91 for TAI-W, and.85 to .91 for TAI-E (Spielberger, 1980). Convergent validity has been shown on the TAI, with the Test Anxiety Scale (Sarason, 1978), the State-Trait Anxiety Inventory (Speilberger et al., 1983).

State Trait Anxiety Inventory (STAI; Speilberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). The STAI consists of two 20-item subscales that aim to differentiate between state and trait anxiety. State anxiety is thought to fluctuate situationally while trait anxiety is a more stable personality trait characterized by tension, apprehension, and increased physiological reactivity. The STAI-state asks people to rate how they are feeling at the present moment, whereas the STAI-trait asks people rate how they feel generally. Higher scores indicate higher degrees of anxiety. Barnes, Harp and Jung (2002) reported that the STAI demonstrated very strong reliability across 117 reliability coefficients, finding an internal reliability coefficient of .91 for the state subscale, and .89 for the trait subscale. The current study only used the STAI-state, as state anxiety is likely to correlate with stereotype threat while taking the WAIS-IV. The STAI has been widely used both in research and in clinical settings and has been reported to have good validity (e.g. Oei, Evans, & Crook, 1990).

The Disengagement Scale (TDS; Major & Schmader, 1998). The TDS is a 12-item measures designed to assess disengagement. A 7-point Likert scale is used, ranging from Strongly Disagree to Strongly Agree. It consists of three independent subscales: discounting, which refers to the validity of standardized intelligence tests, devaluing, which refers which is the degree of importance of doing well in academics, and disengaging, which is the separation of one's self-esteem from performance on intelligence tests. Higher scores indicate higher levels of disengagement. A Cronbach's α value of .75 -.78 have been reported (Lowery & Wout, 2010; Schmader et al., 2001).

RESULTS

Overview

The results of the current study are now presented. Prior to testing the hypotheses of the current study, pre-analyses were conducted to determine whether the two experimental conditions were significantly different from one another, which are presented first. Next, intercorrelations between the main variables of interest are presented. Racial/ethnic differences in the main study variables are then reported, followed by the results of the manipulation checks. Then, results of the a priori hypotheses are reported sequentially. Given that the results of the current study suggested the need for further exploration, post-hoc analyses were conducted, which are reported. Finally, the results of the qualitative interview data are presented. Unstandardized betas are presented throughout, given that this is an applied study.

Pre-Analyses

Condition differences of demographic variables were examined to determine potential confounds. See Table 4 (Appendix R) for χ^2 and ANOVA analyses. There were no condition differences for year in school, racial/ethnic identification, gender, SES, US generational level, college generation level, age, GPA, or ACT score. However, there were marginally significantly differences in income and SAT score. There were more participants in the control condition than in malleability condition who reported that their parents' incomes were over $100,000 (40% vs. 24% respectively) and more students in the malleability condition reported parental incomes between $50,000-75,000 (9% vs.

23%). Across conditions, there were similar levels of parents' income reported for the ranges of under $20,000 (7% vs. 9%), $20,000-50,000 (27% vs. 29%) and $75,001-100,000 (17% vs. 15%). The mean SAT score was 1661.31 for the control condition and 1530.96 for the malleability condition. Due to the fact that income and SES were marginally significant between conditions, they were considered as possible covariates. However, given that 15% of the sample did not report SAT, the sample size would have been reduced to 117, which was less than ideal. Thus, analyses with the full sample of 138 were compared to analyses of the sample of 117, which were determined to be similar. Thus, for the sake of parsimony and a larger sample, analyses without controls are presented.

Intercorrelations between the main variables may be found in Table 3 (Appendix Q). Most of the correlations were in the expected direction. Perceived stereotype threat (PST) was inversely related to three of the four WAIS-IV subtests: Block Design, Digit Span, and Arithmetic. Stereotype vulnerability was inversely related to all WAIS-IV subtests and positively related to ethnic identity (resolution) and disengagement (discounting and disengaging). Implicit theories were inversely associated with the verbal WAIS-IV subtests, such that more entity views were associated with higher test scores (Digit Span, Arithmetic). Implicit theories were also marginally related to PST. Malleability was also positively associated with discounting and disengaging. Post-test test anxiety was related to the verbal WAIS-IV tests (Digit Span, and Arithmetic) but not the performance tests and pre-test test anxiety was only marginally related to Digit Span. Similarly, post-test state anxiety was significantly related to verbal tests (Digit Span,

Arithmetic) but pre-test state anxiety was significantly related to Matrix Reasoning and marginally related to Arithmetic.

Racial/Ethnic Differences among Main Variables

One-way ANOVAs were conducted to analyze racial/ethnic differences on the main variables of interest and contrasts were used to determine locations of significant differences. See Table 2 (Appendix P) for descriptives of the main variables for the sample overall and breakdowns for each racial/ethnic group. For implicit theories of intelligence, Latinos and African Americans were not significantly different from one another but they endorsed significantly more malleable theories of intelligence than did European Americans.

As expected, both racial/ethnic minority groups reported higher levels of ethnic identity exploration and resolution than did European Americans. African Americans and Latino Americans reported similar levels of exploration. There were significant differences in ethnic identity affirmation, with African Americans reporting the lowest level of affirmation. Although Latino Americans reported the highest level of affirmation, they were not significantly different from European Americans. Also as expected, both African Americans and Latinos reported higher levels of stereotype threat, and stereotype vulnerability. African Americans reported significantly higher levels of stereotype vulnerability than Latino Americans who reported significantly higher levels than European Americans.

Although there were no significant racial/ethnic differences for baseline state anxiety, African Americans and Latinos endorsed significantly more state anxiety post-WAIS-IV relative to European Americans. In addition, both racial/ethnic minority groups

endorsed more baseline and more post-WAIS-IV test anxiety than European Americans, but had similar levels to one another.

There were also significant racial/ethnic differences in discounting with African Americans reporting significantly higher levels than European Americans and Latinos, who were similar. There were no differences in African American and European American devaluing, but Latinos devalued significantly less. There were no racial/ethnic differences in disengaging.

Manipulation Checks

The three manipulation checks to test the effectiveness of the malleability condition were examined. There were no condition differences on the ITI. In addition, there were no condition differences for either interview question: "Did you believe you could improve while taking this test?" or "Do you believe that intelligence is a fixed quantity that cannot be changed?" See Table 4 for ANOVA and χ^2 tests. Thus, these findings indicate that the malleability manipulation was not successful. Due to this manipulation failure, the first two experimental hypotheses could not be tested, given that the two conditions were not significantly different from one another.

Hypotheses

The author originally planned to combine the scores of African Americans and Latino Americans to compare racial/ethnic minority persons to European Americans. However, it was determined that the WAIS-IV scores of African Americans and Latino Americans were significantly different on Block Design, t (135) = -3.20, $p < .01$, and Arithmetic, t (135) = -2.28, $p < .01$. Thus, considering the two racial/ethnic minority

groups separately was a more valid approach than collapsing the two stereotype vulnerable groups. Therefore, 3 x 2 ANOVAs were conducted. In addition, it was observed that the distributions of several WAIS-IV subtests were significantly nonnormal, Block Design: *Kolmogorov-Smirnov* (138) = .1, $p < .01$, *Shapiro-Wilk* (138) = .95, $p < .01$) , Matrix Reasoning: *Kolmogorov-Smirnov* (138) = .15, $p < .01$, *Shapiro-Wilk* (138) = .92, $p < .01$, and Arithmetic: *Kolmogorov-Smirnov* (138) = .11, $p < .01$, *Shapiro-Wilk* (138) = .98, $p = .02$, or marginally nonnormal, Digit Span: *Kolmogorov-Smirnov* (138) = .07, $p = .07$. Due to the fact that significant deviations from normality can invalidate results (Howell, 2002), analyses were bootstrapped and replicated 1000 times when the assumption of normality was violated.

Due to the fact that there were no condition differences, it is not possible to examine Hypotheses 3 and 4 within condition as originally intended. Therefore, for these hypotheses, the relationships of the variables of interest were examined for the entire sample, regardless of condition. To test the third hypothesis that ethnic identity would moderate the relationship between PST and WAIS-IV performance, interactional regression analyses were performed. First, all variables in the model were centered on their respective means. Next, interaction terms were created by multiplying PST scores with each subscale of the EIS (exploration, affirmation, and resolution). In hierarchical regressions, EIS subscales and PST scores were entered on the first step, and the interaction term was entered on the second step. There was a significant interaction between PST and EIS-exploration for Block Design (B = -3.18, $t = -2.32$, $p = .02$); the overall model was significant, F (7,130) = 2.05, $p = .05$, and accounted for 11.7% of the variance. This was a small to medium effect ($f^2 = .13$) and the second step did not

significantly improve the fit of the model (R^2 = .18). The interaction suggested that participants with higher ethnic identity appeared to be more vulnerable to stereotype threat, whereas those with lower ethnic identity performed similarly (See Figure 1).

There were no significant interactions for Digit Span, Matrix Reasoning, or Arithmetic. There were no significant interactions between PST and EIS-affirmation for Block Design, Digit Span, Matrix Reasoning or Arithmetic. There were also no significant interactions between PST and EIS-resolution for Block Design, Digit Span, Matrix Reasoning, or Arithmetic. See Table 5 (Appendix S) for details of these analyses.

The fourth hypothesis was that state and test anxiety would mediate the relationship between PST and WAIS-IV performance. Due to a clerical error, PST scores were not collected in the psychology 301 prescreening pool as planned. This prescreening error limits the ability to establish baseline PST, and thus temporal precedence, which makes a true mediational test challenging. PST, state anxiety, and test anxiety were all measured at post-test. Thus, analyses sought to establish statistical mediation, given the limitation of temporal precedence.

The steps recommended by Baron and Kenney (1986) were conducted and tested through a series of regressions. First, an association between the predictor (PST), and the dependent variable (WAIS-IV performance), must be established. Secondly, the association between the predictor and the potential mediator (test/state anxiety), must be established. Thirdly, the mediator must predict the dependent variable in the presence of predictor. Lastly, the relationship between the predictor and dependent variable must drop in significance when the mediator is included in the model.

The first step to test the association between PST and WAIS-IV is the same for test and state anxiety and will be presented once for the sake of simplicity. In four regression models, PST was entered as a predictor and the WAIS-IV subtests were entered as dependent variables. There was no association between PST and Matrix Reasoning (B = -.34, t = -1.21, p = .23), so mediational analyses were not performed for this subtest. PST was significantly related to Block Design (B = -2.49, t = -2.35, p = .02), Digit Span (B = -1.11, t = -2.53, p = .02), and Arithmetic (B = -1.09, t = -4.63, p < .01).

To test the association between test anxiety and WAIS-IV performance, test anxiety was entered as a predictor and the WAIS-IV subtests were entered as dependent variables in four separate regression models. Test anxiety was marginally related to Block Design, (B = -.13, t = -1.89, p = .06), significantly related to Digit Span (B= -2.23, t = -2.76, p = .04) and Arithmetic (B= -.05, t = -2.76, p < .01), satisfying the second criterion of mediation for these relationships.

Next, hierarchical regressions were run to test the mediating role of test anxiety. Test anxiety was entered on the first step and test anxiety and PST were entered in the second step; WAIS-IV subtests were entered as dependent variables in four separate regression models. When test anxiety was controlled (B = -.13, t = -1.98, p = .06), the relationship between PST and Block Design became marginally significant, (B = -2.14, t = -1.84, p = .07). Thus, test anxiety did not mediate the relationship between PST and Block Design. In fact, test anxiety was no longer a significant predictor (B = -.07, t = -1.00, p = .31), when PST included in the model. Similar patterns emerged for the remaining subtests. When test anxiety was controlled to predict Digit Span (B = -.07, t = -2.22, p = .03), the relationship between PST and Digit Span became marginally

significant, (B = -.88, t = -1.83, p = .07), and test anxiety was no longer significant (B = -.07, t = -1.32, p = .19). When test anxiety was controlled to predict Arithmetic (B = -.05, t = -2.76, p < .01), the relationship between PST and Arithmetic remained significant, (B = -.98, t = -4.06, p < .01) and test anxiety became nonsignificant (B = -.02, t = -1.29, p = .21). Thus, test anxiety does not mediate the relationship between PST and Block Design, PST and Digit Span, or PST and Arithmetic. It appears that PST is a stronger predictor of WAIS-IV performance than test anxiety.

Next, the association of state anxiety to WAIS-IV was assessed: state anxiety was entered as a predictor in regression models to predict the four subtests. State anxiety was not associated with Block Design (B = -.07, t = -.75, p = .44) thus state anxiety does not mediate the relationship between PST and Block Design. However, state anxiety was significantly related to Digit Span (B= -.09, t = -2.78, p < .01) and Arithmetic (B= -.07, t = -.76, p = .45). Hierarchical regressions with state anxiety on the first step and state anxiety and PST were entered into the second step of these models; WAIS-IV subtests were the dependent variables. When state anxiety was controlled (B= -.09, t = -2.97, p < .01), the relationship between PST and Digit Span dropped in significance (B= -.89, t = -1.96, p = .05), and state anxiety was still a significant predictor (B= -.07, t = -2.19, p = .03). This pattern suggests that state anxiety partially mediates the relationship between PST and Digit Span. The overall model was significant, F (1, 137) = 4.50, p = .01, and accounted for 6% of the variance (R^2 = .06). The effect was small to medium (f^2 = .06).

When state anxiety was controlled (B= -.05, t = -2.50, p = .01), the relationship between PST and Arithmetic remained significant (B= -1.01, t = -4.12, p < .01) and state

anxiety was no longer a predictor (*B*= -.02, *t* = -1.00, *p* = .20). Thus, state anxiety did not mediate the relationship between PST and Arithmetic.

Given that results suggested that PST could mediate the relationship between test and state anxiety and WAIS-IV performance, exploratory analyses were conducted. Given that the first three steps of mediation have already been established, the final step of mediation is presented here. When PST was entered into the first step of the model (*B* = -2.49, *t* = -2.38, *p* = .02), the relationship between test anxiety and Block Design became nonsignificant, (*B* = -.07, *t* = -.92, *p* = .35) and PST remained a marginally significant predictor, (*B* = -2.14, *t* = -1.82, *p* = .07). The overall model was marginally significant, *F* (1, 137) = 2.82, *p* = .06, and accounted for 4% of the variance (R^2 = .06). The effect was small to medium (f^2 = .04).

When PST was entered into the first step of the model (*B* = -1.11, *t* = -2.53, *p* = .01, the relationship between test anxiety and Digit Span became nonsignificant, (*B* = -.05, *t* = -1.41, *p* = .15) and PST remained a significant predictor, (*B* = -.88, *t* = -1.87, *p* = .06). The overall model was significant, *F*(1, 137) = 3.83, *p* = .03, and accounted for 5% of the variance (R^2 = .05). The effect was small to medium (f^2 = .05).

When PST was entered into the first step of the model (*B* = -1.09, *t* = -5.00, p < .01, the relationship between test anxiety and Arithmetic became nonsignificant, (*B* = -.02, *t* = -1.22, *p* = .22) and PST remained a significant predictor (*B* = -.98, *t* = -4.18, *p* < .01). Thus, results suggest that PST mediates the relationship between test anxiety and WAIS-IV performance. The overall model was significant, *F* (1, 137) = 9.80, *p* < .01, and accounted for 13% of the variance (R^2 = .13). The effect was medium (f^2 = .15).

When PST was entered on the first step (B= -1.11, t = -2.48, p = .02), the relationship between state anxiety and Digit Span dropped in significance (B= -.07, t = -2.19, p = .03), and PST remained a significant predictor, (B = -.89, t = -1.91, p = .06) suggesting that PST partially mediates the relationship between state anxiety and Digit Span. The overall model was significant, F (1, 137) = 4.50, p = .01, and accounted for 6% of the variance (R^2 = .06). The effect was small to medium (f^2 = .06). When PST was entered on the first step (B= -1.09, t = -4.88, p < .01), the relationship between state anxiety and Arithmetic became nonsignificant (B = -.02, t = -1.20, p = .24) and PST remained a significant predictor, (B = -1.01, t = -4.22, p < .01). The overall model was significant, F (1, 137) = 9.75, p < .01, and accounted for 13% of the variance (R^2 = .13). The effect was medium (f^2 = .14). These results suggest that PST mediates the relationship between state anxiety and WAIS-IV performance.

Post-Hoc Analyses

Although the experimental hypotheses could not be tested as originally intended due to the manipulation failure, the data afforded the opportunity to test the hypotheses using alternative methods.

For example, the first two items of the TDS ask about the perceived diagnosticity of standardized and intelligences tests (TDS1: I feel that standardized tests are fair tests of my abilities; TDS2: In general, I feel that standardized tests are a good measure) of my intelligence. Test diagnosticity has been manipulated in previous stereotype threat studies to influence the level of stereotype threat. Presenting tests as diagnostic of ability tends to lead to more stereotype threat than when tests are presented as problem-solving

opportunities that are not used for evaluation (Steele & Aronson, 1995). Given this, students who believe that intelligence tests are indicative of ability may be more susceptible to stereotype threat, and perhaps more sensitive to the malleability prime. Thus, analyses were conducted to determine whether students who endorsed higher beliefs of test diagnosticity were influenced by the malleability manipulation.

To test this diagnosticity hypothesis, TDS1 and TDS2 were combined and averaged to create a diagnosticity index. The sample was divided using a mean split, such that those under the mean were considered to hold non-diagnostic beliefs (N = 97) and those above the mean considered to hold diagnostic beliefs (N = 42). There were no significant condition effects when comparing these two groups for Block Design, F (2, 126) = 1.02, p = .36); Digit Span, F (2, 126) = .40, p = .67); Matrix Reasoning, F (2, 126) = .43, p = .65; or Arithmetic, F (2, 126) = .38, p = .69). Similarly, there were no significant condition effects for PST, F (2,126) = .27, p = .77, TAI, F (2,126) = .63, p = .53, state anxiety, F (2,126) = .63, p = .53, or disengagement, F (2,126) = .33, p = .72. Thus, test diagnosticity beliefs did not influence the effect of the malleability prime.

Another method of assessing variations in the influence of the malleability manipulation is to consider participants' pre-existing implicit theories. It could be that those who already endorsed beliefs in malleability did not experience any additional gain in malleability, whereas those who endorsed entity views could have shifted more. To test this implicit theory hypothesis, the sample was split into malleability and entity theorists. Following the procedures of Dweck et al. (1995), 13 participants with middle-of-the-road scores were dropped from analyses, which resulted in 107 malleability theorists and 19 entity theorists. Given the large difference between the two groups, a

comparison was considered inappropriate and this set of exploratory analyses was discontinued. However, this analysis perhaps gives insight into why the malleability manipulation seemed to be of limited benefit, given that the majority of the sample already endorsed malleable views of intelligence.

The last post-hoc analysis considered participants' endorsement of perceived stereotype threat. Participants who endorsed who scores on the PST reported experiencing more stereotype threat in their daily lives than those with lower scores on the PST. Thus, perhaps those that experiencing higher levels of PST would benefit more from the malleability intervention, relative to those who reported lower levels of PST. To test this perceived stereotype threat hypothesis, a mean split was done to dichotomize the sample into a high PST group (N =54) and a low PST group (N=85). There were no significant condition differences when comparing these two groups for Block Design, F (2, 126) < 0, p = 1); Digit Span, F (2, 126) = .36, p = .70); Matrix Reasoning, F (2, 126) = .79, p = .46; or Arithmetic, F (2, 126) = .46, p = .63). Similarly, there were no significant condition effects for ITI, F (2,126) = .82, p = .44, TAI, F (2,126) = .17, p = .85, or state anxiety, F (2,126) = .85, p = .43. Thus, perceived stereotype threat beliefs did not influence the effect of the malleability prime.

Interview Data

As discussed, the stereotype threat literature has produced mixed findings, leading researchers to debate its existence outside of the laboratory. These findings have been based on quantitative assessments of stereotype threat or implicitly interpreted through patterns of test scores across conditions that vary level of stereotype threat. However, no

studies of which the author is aware directly asked participants about their experience of stereotype threat, which could give further insight into its cognitive and emotional influences. Thus, a qualitative approach was taken to investigate how students perceive and make meaning of the phenomenon of stereotype threat.

Interview data analyses were guided by a phenomenological approach (Giorgi & Giorgi, 2008). Phenomenological psychology aims to illuminate how participants make meaning out of a given experience, or phenomenon. The first step of the phenomenological approach required reading all of the participants' responses to get a sense of the themes that may be present. In the next step, the author and a team of research assistants reviewed the data again to establish coding categories. After establishing the initial coding categories, the responses to the interview questions were coded independently by the research assistants. Coding discrepancies were resolved though discussion facilitated by the principal investigator. The coding scheme was refined as needed, to incorporate themes as they emerged. Results are presented by interview question, describing the emergent categories as well as the percentage of participants that fell into each one. Sample participant answers are provided when illustrative. Codes were allowed to fall into more than one category, as several themes were often present within the same participant response. *Note: examples are taken verbatim, despite grammatical errors.*

Using this phenomenological approach, one goal of the interviews was to examine participants' experience of stereotype threat. Stereotype threat theory would predict that students under stereotype threat would experience more negative emotions, especially those of anxiety and self-doubt, than students in the control condition. Additionally,

stereotype threat would suggest that students would endorse feeling more pressured to perform well under stereotype threat conditions.

To this end, participants were asked 13 open-ended interview questions in order to better understand their opinions about the WAIS-IV, perceptions of malleability, and their experiences of stereotype threat during the experiment and generally at UT. The interview questions can be found in Appendix K. Racial/ethnic and gender breakdowns of results are presented when pertinent. For question one, these breakdowns are incorporated into the text due to the complexity of the question. For ease of interpretation, tables are presented for questions two-thirteen (see Tables 6-17).

Question One: What did you think about the test you just took?

The first interview question was intended to gather information about participants' general impressions of the WAIS-IV. The first question was intentionally broad to allow students to report their most salient thoughts about the test. The coding team identified four categories based on participant responses: difficulty of the test, emotional impact of the test, themes, and opinions/perceptions of the test.

For difficulty of the test, 34% of the sample reported that the WAIS-IV was difficult or challenging ("It was complicated sometimes, kind of hard and confusing."). European Americans (39%) were more likely to report that the test was difficult relative to people of color (28% of African Americans and 30% of Latino Americans). Twenty-four percent of the sample described the difficulty as variable ("It was hard when you don't have paper to do mental math. The word thing was pretty simple."); 4% described it as generally easy ("Simple."), although some of them also reported feeling anxious ("It

was cool -- they were easy, just pressure, like, makes you freak out."); and 38% were neutral and did not spontaneously comment on difficulty ("It was thought-provoking."). Interestingly, European Americans were most likely to spontaneously comment on difficulty of the test (39%), relative to Latino (32%) and African Americans (28%).

Another theme that emerged from the first question was the emotional impact of the WAIS-IV on participants. Although the majority (88%) did not spontaneously report how the test impacted them emotionally, 5% endorsed anxiety ("It was kind of strange, I don't know, it wasn't hard but I felt a little stressful at times."). Latino Americans were the most likely to endorse anxiety (11%), compared to 7% of African Americans and 2% of European Americans. Five percent reported feeling frustrated ("It was challenging but I like it, I like brain games. I like that it kept switching, it would get frustrating than you switch!"). African Americans and European Americans reported similar levels of frustration (7% vs. 8%, respectively), relative to 2% of Latino Americans. Two percent of the sample (2% of European Americans and 4% of Latino Americans) reported feeling unintelligent ("Some of it made me feel dumb - I guess just the math.").

The next category was referred to as themes, which reflected recurring patterns of participant responses. The most frequently identified theme by 16% of the sample was cognitive function/adaptability, which represented having to adapt one's thinking to the situation ("It made me think quickly on my feet."); this category was more often endorsed by European Americans (18%) than people of color (13% of Latinos and 10% of African Americans). Seven percent of the sample gave answers that were slightly defensive, explaining how they might have done better under different circumstances ("I think if I had seen it written down it would have been easier to think through."); this

category was similarly endorsed by all racial/ethnic groups. Five percent of the sample reported feeling like they were unsure of what they were doing ("Some parts were difficult--the numbers one, saying them backwards. Weren't sure what I was doing on some picture part."); 8% of European Americans and 4% of Latino Americans endorsed this category. Four percent of the sample estimated how they did on the WAIS-IV ("Ok I guess - think I might've failed it."); people of color were slightly more slightly to postulate how they performed (5% of Latino Americans and 3.5% of African Americans) compared to 2% of European Americans. Four percent of the sample wondered about the purpose of the test ("I felt like certain patterns and numbers were looking for your ability to see patterns. I'm not sure what the blocks were for. Overall it was interesting test, it wasn't boring or anything."); 8% of European Americans and 4% of Latino Americans fell into this category. Finally, 2% of the sample was surprised about the difficulty of the WAIS-IV ("It was surprisingly hard toward the end of the sections but the beginning was easy."); African Americans and European Americans endorsed this category at similar rates and Latino Americans did not endorse this category.

The last category was participants' perceptions and opinions of the test. Thirty percent reported generally positive opinions of the test ("It was interesting, I had fun."); although some expressed mixed feelings ("It was entertaining I guess, but I felt, the first test with the blocks, at first I wanted to use more blocks, like it was hard for me to visualize. But the other ones were pretty easy."). Thirty-seven percent of European Americans reported positive feelings compared to 31% of African Americans and 23% of Latino Americans. Eleven percent (10% of European Americans, 7% of African Americans, and 5% of Latino Americans) of the sample stated a fact or made an

observation about the test ("It was easy at first, but then it gets harder as it progresses."); 7% reported negative feelings ("I hated it."); and 5% gave ambiguous statements ("It was unique."). Seven percent of the sample reported negative opinions of the test, and this was similar across racial/ethnic groups.

In sum, when asked about their general impressions of the WAIS-IV, participants most often spontaneously commented on its difficulty. The second most common response pattern was giving an opinion or observation about the test. Participants were least likely to spontaneously report how the test affected them emotionally. The theme category described different processes that seemed to occur for participants while undergoing the WAIS-IV. Due to the number of categories identified, the base rates of endorsement were relatively low. Although the highest percentage for a single question one code was 34%, it is important to note that 100% of participant response fell into one of the four identified categories. Thus, low base rates are likely attributed to the number of categories, rather than a poorly defined code.

Question Two: What did you think was the purpose? What did it measure?

In order for stereotype threat to occur, participants must perceive that their intellectual ability is being evaluated. Thus, the second interview question aimed to assess the face validity of the WAIS-IV. The coding team distinguished the concepts of ability, knowledge, and intelligence. Ability was conceived of as a broad category and referenced a wide range of participants' capabilities, from learning to memory to cognitive ability. In contrast, knowledge was conceptualized as a pre-existing fact base. Although the team felt that intelligence was a fairly amorphous construct, for the

purposes of coding, only participants that explicitly mentioned intelligence or IQ fell into this category. The team was resistant to condensing intelligence and ability, arguing that while there is some overlap, they are qualitatively distinct ideas.

Please refer to Table 6 (Appendix T) for categorical and racial/ethnic breakdowns. Fifty-nine percent believed that the test assessed some type of ability. Responses ranged from cognitive ability ("Cognitive function of some sort.") to pattern detection ("To see people's ability to pick up patterns.") to memory ("Memorization. The ability to see something and duplicate it."). European Americans were somewhat more likely than people of color to report that the WAIS-IV measured ability. Twenty-three percent of the sample believed that the purpose of the test was related to speed ("Test people's response times and critical quick thinking I guess."); African Americans were more likely than European and Latino Americans to believe that assessing performance speed was the purpose of the WAIS-IV. Ten percent of the sample identified intelligence as the purpose of the test ("To determine intelligence."); African Americans were less likely than Latino and European Americans to report that intelligence was the purpose of the test. Nine percent of the sample believed that examining performance under pressure was a goal ("To see whether a person tenses up or to see how long it take before they give up."); Latino Americans were less likely than African and European Americans to fall into this category. Six percent believed the goal of the test was to study adaptability ("Abilities to adapt to new tasks."); this category was endorsed by European and Latino Americans, but not African Americans. Latino Americans were the sole respondents for two categories: assessing knowledge was the purpose of the test ("How much you know.") and examining ethnic differences ("If you had different races take the test you

could determine who could do the tests faster. Or how good you are at logical problems."). Five percent of the sample gave vague responses ("I don't know"; "Definitely something with my brain.").

The WAIS-IV may be considered relatively face valid, given that the majority of participants identified examining ability as its purpose. It is interesting to note that although the WAIS-IV protocol has taken measures to reduce anxiety by avoiding the word "intelligence," participants readily perceived that their abilities were being assessed. These findings suggest that the WAIS-IV is perceived as an evaluative situation, which can set the scene for stereotype threat.

Question Three: How did you feel while taking the exam?

According to stereotype threat theory, undergoing an evaluation of one's intellectual ability can provoke negative emotions and thoughts which can be disabling. Thus, the third question aimed to explore the emotional experience of participants undergoing the WAIS-IV to investigate this hypothesis. Please see Table 7 (Appendix U) for categorical and racial/ethnic breakdowns.

Forty-nine percent of the sample reported being emotionally neutral ("Fine."), relaxed ("I felt comfortable, not rushed or anything like that. I felt confident."), or generally calm with occasional negative feelings ("I felt pretty calm. Frustrated at times though."). African Americans were more likely than European or Latino American to report feeling calm during the WAIS-IV. Although most participants were relatively calm during the WAIS-IV, a significant number experienced negative emotions, mostly commonly anxiety. Thirty-eight percent reported negative feelings, such as being

pressured, frustrated, or anxious ("Nervous, like rushed."; "Frustrated, I wanted to give up."). European Americans were more likely than people of color to report negative feelings about the test. Fifteen percent noted that they became more nervous as the difficulty increased ("I started out pretty confident but as the questions got harder I got more nervous, especially after that one block I couldn't figure out."). People of color were much more likely than European Americans to report that their anxiety increased in relation to the difficulty of the items. Six percent reported feeling stupid ("A little stupid-during the math, anxious."); African Americans were slightly less likely to report feeing stupid relative to European and Latino Americans. Four percent of Latino Americans reported feeling apathetic or bored, ("I got more apathetic as it progressed."). Six percent of the sample gave ambiguous statements that were difficult to interpret emotionally ("I wanted it to be over."; "challenged").

Question Four: While taking the test, did you believe you could improve your performance?

Question four served as one of the manipulation checks and aimed to assess participants' perceptions of malleability. Please see Table 8 (Appendix V) for categorical and racial/ethnic breakdowns. The majority of the sample (65%) generally agreed with the statement with a straightforward yes. People of color were more likely to agree with this malleability statement relative to European Americans. Eight percent of the sample gave a qualified yes, specifying that performance could improve on certain tests ("On some of them, because they're longer and I can learn what they're doing, figure out the patterns."). Latino Americans were slightly more likely than European and African

Americans to give a qualified response. Five percent gave tentative answers ("I guess so;" "maybe"), and levels of endorsement were similar across racial/ethnic groups. Nineteen percent did not believe they could improve their performance on the test. These findings are consistent with those of Dweck et al. (1995) who argued that an incremental view of intelligence seems to be more palatable to most people. Results also suggested that the manipulation did not succeed in shifting participants' implicit theories. European Americans were most likely to agree with the entity theory, followed by African Americans, then Latino Americans.

Question Five: How effective do you think the test you were given was in determining your ability as a university student?

Given that WAIS-IV scores have been used for educational placement purposes, the current research wanted to consider student opinion of this practice. Please see Table 9 (Appendix W) for categorical and racial/ethnic breakdowns. Fifty-three percent believed that it was irrelevant ("Poor, because being a good student isn't necessarily measured by tests like this."), and members of the three racial/ethnic groups reported this equally. Twenty-nine percent of the sample believed the WAIS-IV could be somewhat effective ("I would say not incredibly effective, doesn't measure a studying factor, but does take quick reasoning into account."); this perception was endorsed most often by Latino Americans, followed by European Americans, then African Americans. Fifteen percent of the sample believed that it was effective ("I thought it was alright at it because some of it was challenging. I guess you have to be able to do stuff like that to be successful."). African Americans were the least likely to believe that the WAIS-IV was

an effective determinant of one's academic ability. In sum, the majority of the sample did not feel that the WAIS effectively assessed their ability as a university student.

Question Six: Have there been times when you felt you were negatively stereotyped regarding your intelligence on this campus? Prior to coming to UT?

The sixth question aimed to assess the perceived frequency of being intellectually stereotyped. The literature suggests that racial/ethnic minority participants would be more likely to endorse being intellectually stereotyped than non-minority participants, which was supported. Please see Table 10 (Appendix X) for categorical and racial/ethnic breakdowns. About half of respondents endorsed having their intelligence stereotyped (49%; 67 people), and half did not (50%; 69 people). Thirty-two percent reported experiencing discrimination at UT and 32% reported that they had experienced discrimination prior to coming to UT. Some participants reported that they had never experienced discrimination before UT. Others reported that they experienced discrimination prior to UT but had not since they arrived at UT. Thus, despite the identical percentages, these experiences were endorsed by different participants.). As expected, African and Latino Americans reported higher incidence of discrimination at UT, although a significant number of European Americans also reported experiencing discrimination. Interestingly, European Americans were most likely to report experiencing discrimination prior to UT.

Participants gave various reasons as to why they experienced discrimination, most commonly related to ability (37%): people doubting their intelligence ("Yes, athletes are said to not be as smart, along with blacks not being as smart as the majority of the school.

No (to prior)".); their presence at UT ("Not that I can think of; Yes, when people were shocked that I got into UT."); and to a lesser extent being ignored by teachers ("Yes, um, well, sometimes I feel like some professors tend to pay attention to other students who aren't of my ethnicity. No to prior."); and assumptions of laziness ("No; yeah I always hear Mexicans are supposed to be lazy, not expected to go to college."). All other answers did not fall into the above themes or were vague ("Yeah, sometimes. The school I came from is recognized as lower-class."; "No I don't think so; yeah just because I'm a girl."). Thus, having one's intelligence stereotyped was a fairly common experience for students in this sample and seemed to occur at similar rates both before and after coming to UT.

Question Seven: Do you believe there are stereotypes for your ethnic group regarding your intelligence?

Question seven was meant to assess participants' awareness of racial/ethnic stereotypes. Please see Table 11 (Appendix Y) for categorical and racial/ethnic breakdowns. Seventy-eight percent endorsed that there were ethnically based stereotypes. Nearly all people of color acknowledged the existence of such stereotypes, relative to half of European Americans. Forty-one percent of the sample endorsed stereotypes related to ability and intelligence. Members of all three racial/ethnic groups endorsed this question: ("Yes. Most people think Hispanics aren't as smart as like Caucasians or Oriental people."; "Yes, um, well like black people aren't as smart as other people, black people are ghetto, loud, like to fight. Those are the main stereotypes."; "Yes. That white people are smart."). On the other hand, 19% said there were no stereotypes for their racial/ethnic

group. Some participants of these participants acknowledged stereotypes of other groups ("Not mine personally. Asians and Indians are supposed to be smart; We're average - I guess the only ones without stereotypes."). Two percent were unsure ("I'm not sure. It's more like that for other ethnicities."). Interestingly, 25% referenced cultural stereotypes such as teen pregnancy and employment ("Yes. That we're not smart and we just belong working in a kitchen.") Seven percent referenced stereotypes related to motivation ("Yes - maybe that we're not focused, maybe we don't take school seriously."). Thus, the majority of the sample acknowledged common stereotypes for each major racial/ethnic group within the current study. This is consistent with previous findings that societal stereotypes are pervasive and common knowledge (Steele & Aronson, 1995).

Question Eight: Do you believe there are stereotypes for your gender regarding your intelligence?

Question eight was meant to assess participants' awareness of gender stereotypes. Please see Table 12 (Appendix Z) for categorical and racial/ethnic breakdowns. Interestingly, 43% reported that there were no gender stereotypes, suggesting that racial/ethnic stereotypes may be more salient than gender stereotypes. African Americans were most likely to report the existence of gender-based stereotypes, followed by European Americans, then Latino Americans. In the sample generally, females were more likely than males to reported gender-based stereotypes. Forty-nine percent of the sample reported stereotypes related to intelligence, including the common stereotype about men being superior at math and science relative to women ("Yes, that we're not as good at math or science as males.") but also including less common stereotypes ("Yes, I

think people think guys are somewhat more lazy and not as hard-working."). Ten percent referenced gender stereotypic roles ("Yes, women should stay at home and not think about going to school.") or gender dominance in employment ("Maybe. Like if you're looking for a job, like some engineering stuff sometimes they hire men.").

Question Nine: Have you ever had a time that you were so worried about appearing stereotypical that it interfered with what you were trying to accomplish? Do you think that happened today?

Question nine asked participants indirectly if they have experienced stereotype threat. Please see Table 13 (Appendix AA) for categorical and racial/ethnic breakdowns. The majority of the sample (85%) said no; European and Latino Americans were more likely to deny experiencing stereotype threat than African Americans. Ten percent of the sample endorsed that they had a time when they worried about appearing stereotypical, but not while taking the WAIS-IV ("Yes, coming here [UT] at first. When I got here it was like a freaking culture shock. It's so diverse here. I didn't want to be the stereotypical black girl."); there were no racial/ethnic difference in endorsement of this experience. However, 4% of the sample reported experiencing stereotype threat during the WAIS-IV ("Probably. Yes, maybe."). African Americans were the most likely to report stereotype threat during the WAIS-IV.

It is interesting to note that many of the participants who endorsed experiencing stereotype threat during the WAIS-IV were often tentative and unsure of their answers, which may be consistent with Steele's conceptualization of stereotype threat as a nebulous "threat in the air" (Steele, 1997). Given that the PST measure significantly

predicted many of the WAIS-IV subtests, it may be argued that while stereotype threat seems to be related to test performance, it is largely an unconscious phenomenon as Steele suggested.

Question Ten: Intelligence has sometimes been described as something that you are born with that does not change, meaning that however much you have when you are young stays the same for the rest of your life. Do you agree?

Question ten also served as a manipulation check, meant to assess level of agreement with entity theory. Please see Table 14 (Appendix AB) for categorical and racial/ethnic breakdowns. The majority of the sample (75%) disagreed ("No. I think your intelligence can get better or worse. It's all about exercising your brain I guess."). People of color were more like to disagree with the entity theory relative to European Americans. The preponderance of the incremental view of intelligence is consistent with previous literature (Dweck et al., 1995). However, 14% percent of the sample expressed mixed agreement and qualified their answers ("To a certain extent yea. But I think if you practiced you could increase your IQ a little but there is probably a predetermined range."). Eleven percent agreed with the entity view of intelligence, ("I do think that you are born smart or dumb."). Interestingly, African Americans were a bit more likely to endorse the entity theory, compared to Latino and European Americans.

Question Eleven: Have you ever taken the WAIS? Have you been exposed to it in any class?

Question eleven asked participants about their previous knowledge of the WAIS-IV. Please see Table 15 (Appendix AC) for categorical and racial/ethnic breakdowns. The

majority of the sample (75%) had never heard of it. Latino and European Americans were somewhat more likely to report that they had not heard of the WAIS-IV compared to African Americans. Eighteen percent had heard about the WAIS-IV, most often through introductory psychology ("Yea, in psy 301.") this percentage was similar across racial/ethnic groups. Five percent reported that they had taken similar tests as children, ("I think I took something similar a long time ago when I was a kid."), which was also similar across racial/ethnic groups.

Question Twelve: Do you think you've ever experienced the feelings that you just read about, called stereotype threat? Please explain.

Question twelve asked participants directly if they have ever experienced stereotype threat after reading a description of it in the debriefing form. Please see Table 16 (Appendix AD) for categorical and racial/ethnic breakdowns. The majority of the sample (65%) simply said no. Some participants reported that they had not experienced stereotype threat due to the racial/ethnic homogeneity of their background ("Not really. Where I'm from it's all one race, it's all Mexican and stuff."). Others did not feel stereotype threat was applicable to them ("No, don't believe being white will make my score lower in any way."). However, it is interesting to that a higher percentage endorsed the question after reading a description of stereotype threat, in contrast to being asked in question nine. Thirty-two percent said they had experienced stereotype threat. ("Yes, like because I'm Hispanic."). African Americans were the most likely to report experiencing stereotype threat, followed by Latino Americans then European Americans. Of those that endorsed the item, 8% reported the feeling during tests "yes, taking the SAT"), 10%

described experienced of being affected by stereotypes generally ("Yeah, I know people have thought stereotypes toward me but they haven't really ever said it."), and 5% described feeling prepared or disproving such stereotypes ("I've heard it, I tend to misprove it.") Some agreed with the feeling of stereotype threat, but did not identify specific examples ("Very rarely, sometimes. I don't really have an example, there are just random things you notice. It does suck."). Descriptions such as these as are reminiscent of racial microaggressions (Sue et al., 2007), in that victims are left with a vague feeling of insult but being unsure if there was a perpetration or not.

Question Thirteen: Do you think you experienced stereotype threat in today's study?

Question thirteen directly asked participants if they felt they had experienced stereotype threat during the WAIS-IV. Please see Table 17 (Appendix AE) for categorical and racial/ethnic breakdowns. The majority of the sample (95%) said no without elaboration and 5% said yes ("Kind of yeah, I didn't know that this was testing or had anything to do with my race, but it's always in the back of my mind. Oh I have to do good otherwise they'll think all Hispanics are dumb."). African Americans were somewhat more likely than European Americans and Latino Americans to report experiencing stereotype threat during the study.

Thus, the rate of endorsing stereotype threat experienced during the WAIS-IV appears to be fairly similar whether asked directly or indirectly (85% to 95% respectively). Again it is interesting to note this conscious experience of stereotype threat in contrast to relationship found between perceived stereotype threat and WAIS-IV test scores observed quantitatively.

DISCUSSION

Malleability and Stereotype Threat

The first hypothesis that racial/ethnic minority groups would experience less stereotype threat in the malleability condition compared to the control condition and the second hypothesis that members of racial/ethnic minority groups would perform better in the malleability condition could not be tested due to the manipulation failure. Manipulation checks showed no condition differences, suggesting that the experimental manipulation was unsuccessful. Participants in the control condition were no more likely to endorse an entity theory than participants in the malleability condition. Post-hoc analyses aimed to test the experimental hypotheses in alternative ways also failed to find differences between the two conditions. However, one post-hoc analysis indicated that the vast majority of the sample already held incremental theories. Thus, it may be that participants were not affected by the malleability prime because of their pre-existing views.

In addition, two interview questions served as manipulation checks and provide further insight. The majority of participants (73%) believed they could improve on at least some of the WAIS-IV subtests and most (75%) disagreed with the entity theory of intelligence. Thus, the majority of the sample endorsed incremental beliefs of intelligence. Humans have a natural tendency to seek for information that enhances their self-esteem (Rydell et al., 2009; Rydell & Boucher, 2010). Furthermore, it has been shown that self-affirmation can buffer against stereotype threat (Martens et al., 2006). Thinking of abilities as malleable may protect self-esteem by instilling a sense of hope

that performance and skills may be improved through increased effort. While Dweck et al. (1995) noted that incremental theory is inherently more appealing than entity theory for many people, it seems that this tendency also has cultural elements. Both quantitative and qualitative data revealed that students of color were more likely to endorse more malleable theories of intelligence than European Americans. In a society with many disparaging stereotypes about people of color, African Americans and Latinos may hold malleable theories in order to protect their self-esteem.

The weakness of the experimental manipulation makes drawing conclusions about the present study with regard to stereotype threat difficult. Although the control condition was expected to have a baseline level of stereotype threat thought to be endemic to evaluative situations (Brown & Day, 2006; Davis et al., 2006; Steele & Aronson, 1995; Wout et al., 2009), the current study neither supported nor refuted this hypothesis. The presence of stereotype threat during the WAIS-IV administration can neither be directly confirmed nor denied.

A comparison of the quantitative and qualitative findings proves useful in illuminating the experience of stereotype threat in the study. On the one hand, the majority of students denied experiencing stereotype threat during the WAIS-IV whether asked indirectly (85%) or directly (95%), which suggests that stereotype threat may not interfere with performance. However, it is interesting to note that many of the participants who did endorse experiencing stereotype threat while being tested were often tentative and unsure of their answers ("Probably; I guess"), which may be consistent with the conceptualization of stereotype threat as a nebulous "threat in the air" (Steele, 1997). Participants were more likely to report general feelings of stereotype threat, giving

examples of being stereotyped in social situations or classrooms, rather than in the immediate testing environment. Participants were also more likely to report stereotype threat when asked directly, with 32% endorsing stereotype threat after reading a description of the phenomenon. Interestingly when asked directly about stereotype threat, African Americans were less likely to report that they had ever experienced it, relative to Latino and European Americans. African Americans were also more likely than Latino and European Americans to believe that the purpose of the WAIS-IV was to study speed, and less likely to report that the purpose was to study ability or intelligence than the other two groups. Perhaps conceptualizing the test as a measure of speed was less threatening for African Americans and contributed to less perceived stereotype threat. In addition, both African and Latino Americans were less likely to comment on the difficulty of the WAIS-IV, although people of color were more likely than European Americans to report that they became more anxious when the difficulty of the items increased. European Americans were more likely to give more affective responses than people of color, expressing more positive or negative comments about the WAIS-IV. These findings suggest that people of color may be apt to minimize or withhold their feelings in regards to test performance in an effort to preserve their self-esteem.

On the other hand, while participants of color tended to deny that they experienced stereotype threat during the WAIS-IV, they endorsed higher levels on the quantitative measure of perceived stereotype threat (PST) and stereotype vulnerability than European Americans, which is consistent with previous literature (Mendoza-Denton et al., 2002). The mean level on the PST for African Americans and Latino Americans was moderate, indicating some endorsement of stereotype threat, which is in contrast to

the verbal denial of stereotype threat. Despite the fairly modest levels of PST, PST scores were predictive of WAIS-IV scores, such that participants with higher levels of PST tended to have lower WAIS-IV scores. These findings suggest that students of color may not have experienced stereotype threat in the immediate testing situation. However, the *perception* of the existence of stereotype threat in society more generally, as measured by the PST, is related to performance on the WAIS-IV.

These findings reflect the prevalence of negative stereotypes of people of color within American society (Steele & Aronson, 1995). Roughly half of the sample reported that they had been stereotyped regarding their intelligence, and of those, 32% endorsed being stereotyped on campus. Consistent with expectations, African Americans reported higher awareness of ethnic-based stereotypes as well as the highest rates of discrimination at UT, followed by Latino Americans, and then European Americans. Interestingly, there were no racial/ethnic differences in discrimination prior to UT. It is also notable that half of the sample did not report being stereotyped. Seventy-eight percent of the sample endorsed being aware of ethnically-based stereotypes of intelligence. While most participants reflected the common stereotype of women having difficulty with math, a surprising 43% of the sample reported no gender stereotypes, suggesting that racial/ethnic stereotypes are more salient. Wout et al. (2009) noted that African Americans are more aware of negative stereotypes than are women. In addition, although both African Americans and Latino Americans can be vulnerable to stereotypes, African Americans reported more perceived stereotype threat than their Latino counterparts and nearly twice as much stereotype vulnerability. In addition, although the percentage of students reporting that they experienced stereotype threat during the

WAIS-IV was small, African Americans were most likely to report that they had, when asked indirectly or directly, compared to Latino and European Americans. The current study suggests that African Americans are especially vulnerable to the effects of negative stereotypes. It is interesting to note that African Americans, whom were the most vulnerable to stereotype threat in the current study, also scored the lowest on the WAIS-IV subtest scores. Perhaps greater efforts are needed to reduce stereotype vulnerability and stereotype threat for African American college students.

Interestingly, Latino Americans were somewhat less likely than the other two racial/ethnic groups to report awareness of gender-based stereotypes. It is interesting to consider how the values of *marianismo* and *machismo* may be related to these findings (Arredondo & Perez, 2003). Marianismo and machismo are culturally prescribed expectations of how women and men, respectively, should behave in many traditional Latin American cultures. The ideal woman from the perspective of marianismo is selfless, devoted to her family, and nurturing; while the ideal man from the perspective of machismo is strong, protective, and authoritative. One might argue that gender roles are deeply instilled in Latino culture and perhaps less likely to be perceived as stereotypes.

To make sense of the study's findings, differences in assessment of stereotype threat via the PST and interview should also be considered. The process of responding to an open-ended question is quite different than indicating degree of agreement to designated items on a Likert scale. Perhaps participants were less likely to endorse stereotype threat when being interviewed by European American researchers, but were more comfortable endorsing such feelings on paper. In addition, semantic differences in wording could have also influenced participants' responses to questions. Given that the

PST measure significantly predicted many of the WAIS-IV subtests, it may be argued that while stereotype threat seems to be related to test performance, it is an unconscious phenomenon, which is consistent with Steele's (1997) theory. That is, while students recognize the presence of stereotype threat in society, they may be more reluctant to endorse it in a given situation.

It has been argued that when the presence of stereotype threat is unclear, there is a motivation to discount its presence to avoid contending with the emotional consequences of negative stereotypes (Steele et al., 2002; Wout et al., 2009). Since stereotype threat was not directly induced in the current study, the testing situation was fairly ambiguous, leaving students to make their own interpretations regarding stereotype threat. Thus, in the absence of blatant threat, students could have minimized its presence. This may also explain why several studies have failed to find evidence of stereotype threat in real-world settings (Cullen et al., 2004; Sticker & Ward, 2004).

Central to the theory of stereotype threat is that examinees experience apprehension about being evaluated on the basis of their race in a situation where they may be judged negatively (Steele & Aronson, 1995; Steele, 1997). Indeed, results of the current study reflected a pattern of evaluation apprehension. A significant number (almost 40%) of participants reported negative feelings (e.g. frustrated, pressured, anxious, stupid), especially in response to perceived difficulty, and the majority of students in the sample (58%) reported that at least a few of the WAIS-IV subtests were challenging. Results suggest that these negative feelings in response to evaluative pressure could have made students vulnerable to stereotype threat. However, participants did not readily link the source of this anxiety to racial/ethnic stereotypes.

Perhaps the original intention of the study to reduce the naturalistic level of stereotype threat by enhancing malleability was not realized due to lack of perceived threat in the immediate testing environment. In the current sample, mean scores of PST indicated low endorsement for European Americans, neutral attitudes for Latino Americans, and slight agreement for African Americans. Thus, perceived stereotype threat was fairly low for these students. The failure of the PST mean split to reveal condition differences is also suggestive of a possible ceiling effect on the potency of the malleability condition. Perhaps for individuals high in perceived stereotype threat, the emphasis on malleability might have provided encouragement in the face of threat to one's ability. However, perhaps because the sample was generally low in stereotype threat, they did not benefit from reading about malleability because they did not doubt their abilities.

Another possibility for the lack of condition differences is that the manipulation may have been too brief and subtle to change participants' implicit theories of intelligence. Aronson et al. (2001) and Inzlicht et al. (2003) used longer-term and more extensive interventions to increase malleability such as videos about neuroplasticity or having participants undergo writing sessions to indoctrinate themselves with theories of malleability over several weeks. Although Burns and Isbell (2007) commented on the amenability of implicit theories to change, the current study suggests that reading a paragraph about malleability is not sufficient to affect participants' beliefs. Furthermore, the manipulation may not have been credible. In a study examining the role of disengagement in intellectual test performance, Major, Spencer, Schmader, Wolfe and Crocker (1998) found that the oft-utilized race prime manipulation failed to affect the

self-esteem levels of African American college students. Major et al. suggested that African Americans did not believe that the test was culturally unbiased as it was purported to be. It could be that the manipulation of the current study was not strong enough to override pre-existing stereotype threat. Aronson et al. (2002) and Good, Aronson, and Inzlicht (2003) argued that malleability interventions may not change perceptions of stereotype threat, but instead function by changing students' behavioral responses. In the future, similar studies could explore this idea by asking participants if they believed the test they had just taken was biased.

Ethnic Identity and WAIS-IV Performance

In addition to experimental hypotheses, the current study sought to provide insight regarding the role of racial/ethnic identification on test performance. The third hypothesis that perceived stereotype threat would moderate the relationship between ethnic identity and WAIS-IV scores received mixed support. Perceived stereotype threat significantly interacted with ethnic identity (affirmation) to predict Block Design scores, but no other subtest scores (see Figure in Appendix AF).

It appears that at low levels of PST, ethnic identity does not have an influence of Block Design scores. However, as PST increases, higher ethnic identity is associated with lower Block Design scores. These findings are consistent with previous literature that has found that higher racial/ethnic identity is associated with greater vulnerability (Armenta, 2010; Davis et al., 2006; Ployhart et al., 2003), which is consistent with the theory of stereotype threat. Greater identification with one's racial/ethnic group renders stereotypes more relevant and thus more potentially threatening. However, it is unclear why this

interaction occurred on Block Design but none of the other WAIS-IV subtests. For the majority of the WAIS-IV subtests, ethnic identity did not interact with perceived stereotype threat and stereotype vulnerability. Thus, the effect of racial/ethnic identity may be relatively limited in the context of intellectual performance. Previous researchers (Awad, 2007, Sellers et al., 1998) have asserted that ethnic identity is relatively distal in the explanation of academic performance and that other factors, such as academic self-concept, consequences of school failure, and the value of education, are more relevant. In the current study, it is possible that more proximal factors such as test anxiety and previous experience with tests are more predictive of WAIS-IV scores, which may contribute to the lack of significant results.

In the present study, correlations revealed significant racial/ethnic differences in psychological disengagement, with African Americans reporting the highest levels of discounting, followed by European Americans, and Latinos. This is consistent with Major et al. (2002), who suggested that African Americans chronically disengage their self-esteem from their performance on intelligence tests. In the current study, discounting was inversely related to Block Design, Matrix Reasoning, and Arithmetic subtests, and marginally associated with Digit Span, which indicates that those scoring lower tend to discount the feedback of such tests. By discounting the validity of tests to accurately reflect ability, poorer performance is less threatening to self-worth. Interestingly, Latinos reported similar levels of discounting to European Americans and indicated lower levels of devaluing than both other racial/ethnic groups. This finding is consistent with previous literature that education is highly valued in Latino culture (Hill & Torres, 2010).

Inconsistent with previous literature (Major & Schmader, 2003) was the lack of difference observed between African Americans and European Americans on devaluing and disengagement. It is important to note that overall levels of devaluing were low, indicating that success on academic and intellectual tests were important overall in the sample. However, more research is needed to understand the gap between valuing education and academic performance (Hill & Torres, 2010), given that graduation rates for African American and Latino American students tend to be lower than that of European Americans.

Anxiety and Stereotype Threat

The fourth hypothesis that test and state anxiety would mediate the relationship between PST and WAIS-IV performance was not supported. PST was more strongly related to WAIS-IV scores than both state and test anxiety, which suggests that anxiety may not be the mechanism by which stereotype threat works. However, it is possible that mediation occurs in the other direction, such that stereotype threat is the mechanism by which anxiety has an effect on test performance (Jordan & Lovett, 2007). In the current study, exploratory analyses suggested that perceived stereotype threat mediated the relationship between test and state anxiety and the verbal WAIS-IV subtests (Digit Span and Arithmetic). However, this finding is tentative due to the lack of temporal precedence, given that PST was measured after test/state anxiety. Thus, future studies need to establish temporal precedence to test this possibility.

The current study is consistent with Osborne (2001), who found that anxiety accounts for significant portion of racial/ethnic differences in achievement test

differences. There were significant racial/ethnic correlational differences regarding anxiety. Although there were no significant racial/ethnic differences for pre-test state anxiety, African Americans and Latinos endorsed significantly more state anxiety post-test relative to European Americans. In addition, both racial/ethnic minority groups endorsed more pre-test and more post- test anxiety. Thus, it appears that the experience of evaluation evokes more anxiety for racial/ethnic minorities than their European American counterparts. Furthermore, there were significant positive correlations between perceived stereotype threat and both pre- and post- test and pre- and post- state anxiety, suggesting stereotype threat is associated with evaluative apprehension, which is consistent with Steele and Aronson's (1995) assertion that anxiety arises in evaluative situations for stereotype threatened groups.

Conclusions

Despite the manipulation failure of the current study, findings suggest that perceived stereotype threat and stereotype vulnerability are at least indirectly associated with intellectual test performance, since it was found that they were related to most of the WAIS-IV subtests (with the exception of Matrix Reasoning). Future research should continue to compare the immediate perception of stereotype threat within evaluative situations as compared to the perception of stereotype threat in general societal situations. Furthermore, there is a rich literature that supports the importance of stereotype threat on academic performance. The current study suggests that stereotype vulnerability should also be considered in interventions aimed to improve the academic and intellectual performance of people of color. Aronson and Inzlicht (2004) found that African

Americans high in stereotype vulnerability were less accurate in predicting their academic performance. In addition, their academic self-esteem was more variable than that of European Americans and African Americans low in stereotype vulnerability. This suggests that stereotype vulnerability might contribute to self-doubt and lower accuracy of predicted performance, which could in turn have negative ramifications upon test scores. Thus, future researchers might consider the roles of both stereotype vulnerability and stereotype threat when developing interventions focused on enhancing academic self concept and self efficacy; two variables that are likely related to intellectual test performance.

The results of the current study have several implications for researchers, educators, and clinicians providing intellectual assessment. Controversy continues to surround the continued use of the WAIS-IV for admission into special education and gifted programs and the pattern of racial/ethnic differences. Although examining the causal factors of such differences was beyond the scope of the current study, it is important to keep in mind that the reasons for racial/ethnic differences on intelligence tests continue to be poorly understood. This is highlighted by the fact that intraracial differences continues to be larger than interracial differences (Loehlin, 2000; Neisser et al., 1996). As several researchers have emphasized (Goldstein, Scherbaum, & Yusko, 2010, Neisser et al., 1996), it is important that researchers avoid holding a "case-closed" mentality regarding the development and assessment on intelligence.

As noted due to persistent racial/ethnic differences, much debate still concerns the use of cognitive ability assessment. Interestingly, over half of the sample believed that the WAIS-IV was largely irrelevant in determining their abilities as a university student.

Participants felt that factors such as studying were not captured by the WAIS-IV. The students' critiques echo arguments in the literature that intellectual tests only moderately predict actual university student performance (Williams, 1996), although they are routinely used for academic admission and placement.

Administrators of special education programs should also consider the possible effects of stereotype threat and anxiety when considering applicants. Furthermore, psychologists providing assessment for racial/ethnic minority examinees should be aware that they may experience more test and state anxiety relative to European Americans when undergoing the WAIS-IV. Also, because the verdict is still out on if and how stereotype threat affects performance in real-world settings, test interpretations could consider the potential role of stereotype threat. In addition, graduate training programs would to do well to include education about stereotype threat in assessment courses in which the WAIS-IV is learned.

The current study has several limitations to note, most obviously the weakness of the malleability manipulation. Secondly, the error during pre-screening limited the current study's ability to detect change in implicit attitudes and perceived stereotype threat that may have occurred due to the manipulation. Furthermore, the ability to test mediation was limited by the inability to establish temporal precedence in the relationship of perceived stereotype threat and anxiety. Another limitation included the relatively small number of African Americans (N=29), compared to Latino Americans (N = 57) and European Americans (N = 52). Lastly, the study focused on college students who participated in the study for research credit. Thus, the stakes may not have been high enough to elicit stereotype threat, given that there were no negative consequences of

performance in this context, which is in sharp contrast to the life-changing impact that real-world assessment results can have. Regarding previous exposure to the WAIS-IV, few college students had heard of the test and even fewer had taken it. College students are a generally high functioning and high performing group that has already demonstrated academic competence. Thus, the WAIS-IV may be not relevant to this population.

Thus, a more ecologically valid study would focus on participants likely to encounter the WAIS-IV, such as applicants for disability services or special and gifted education programs. Attaching the real-world value of these programs could enhance motivation to succeed and may provide more accurate results. Jordan and Lovett (2007) noted that it is unknown whether increasing the stakes would exacerbate stereotype threat by increasing the pressure of the situations or ameliorate stereotype threat by increasing motivation. However, Steele and Davies (2003) would predict that increasing the stakes would augment stereotype threat by increasing the pressure of the situation.

Despite these limitations, the study also contributed to the limited literature of stereotype threat, ethnic identity, and WAIS-IV performance. To the author's knowledge, this is the first study to examine the relationship of perceived stereotype threat to intellectual test performances on the WAIS-IV. Also unique were the efforts to increase external validity within a laboratory setting by examining naturalistic levels of perceived stereotype threat without explicit stereotype threat induction. Past stereotype threat studies have been criticized for creating artificial testing situations (Stricker & Ward, 2008), only demonstrating that when stereotype threat is primed that it can have deleterious effects. In addition, the current study also contributed to the stereotype threat literature by expanding the cross-cultural research typically done. Most studies have

compared two racial/ethnic groups (most often African Americans and European Americans). However, the present study examined the majority group as well as two racial/ethnic minority groups that have been stereotyped to have low intelligence. Another contribution was the qualitative data that gave insight into examinees' subjective experiences of stereotype threat after undergoing intellectual test performance, which had not been previously done. These answers provide insight into how examinees feel when presented with the WAIS-IV, which is important in understanding factors that contribute to intellectual test performance.

Much could be learned if the present study were to be replicated, with some important revisions. Stronger efforts to detect the presence of stereotype threat in the immediate testing environment should be included, such as a baseline measure of perceived stereotype threat. A condition in which stereotype threat was primed explicitly as is typically done in the literature could be included in order to provide a control group. Pre-test and post-test measures of disengagement could serve to examine whether stereotype threat conditions increase the likelihood of disengagement for threatened participants. Baseline measures of implicit theories would also be helpful to determine the effectiveness of the manipulation. Including other self perception measures such as self-esteem and academic self concept would be useful to determine how evaluative situations can affect the identity of threatened students. In addition, the malleability manipulation would need to more extensive in terms of length and content, administered over the course of several sessions. Future studies should examine whether more extensive, longitudinal interventions focusing on reducing stereotype threat and

stereotype vulnerability can affect performance on the WAIS-IV and other tests of intellectual ability.

Other research directions include considering alternative methods of reducing stereotype threat in addition to malleability. For example, examining whether interventions focused on reducing test anxiety for examinees of color could potentially yield useful information. Also important is continued research regarding the mechanisms and moderators of stereotype threat. Perhaps most importantly, more real-world studies of stereotype threat are needed to determine to what extent the phenomenon occurs outside of the laboratory settings in which it has been replicated numerous times.

The current study speaks to the methodological difficulties of translational research of stereotype threat. In light of the continuing debates about the existence of stereotype threat in the literature, some researchers may question if it is possible to properly measure such an amorphous concept as stereotype threat in real-life settings. As Steele (1997) has argued, stereotype threat may exist on an unconscious level, making it elusive to quantify and observe in actual testing situations. Yet as several authors (Jordan & Lovett, 2007; Steele & Davies, 2003) have noted, the stereotype threat literature has produced consistent and robust findings that merit the attention of practitioners and researchers. However, far more studies in real-world settings are needed to test the ecological validity of the stereotype threat theory.

Appendix A-Condition Scripts

Malleability prime:

Thank you for agreeing to participate in this study. Today we will be asking you to complete a series of exercises, which the experimenter will guide you through momentarily. These exercises measure abilities that tend to improve with practice. Think of the mind as a muscle: the more you use it, the stronger it becomes. When in a novel situation, the brain can form new connections between neurons, which allows you to process information and understand it in a different way. This ability is known as plasticity or malleability. Today, there may be times that you encounter difficulty, but remember that your brain can adapt to the challenge.

Control script:

Thank you for agreeing to participate in this study. Today we will be asking you to complete a series of exercises, which the experimenter will guide you through momentarily. The study length varies from 60-90 minutes, but usually does not take longer. After the exercises, you will complete an interview, followed by a questionnaire packet. Then, after you have completed all part of the study, we will give you research credit. Again, we appreciate your time.

Appendix B-Stereotype Activation Measure

Please complete the following word fragments with the first thoughts that come to your mind.

1. __ __ C E
2. F L __ __ __ __
3. S U __ D __ __
4. M O __ __ E
5. L A __ __
6. __ __ A Y
7. __ O __ __ E
8. __ __ L D
9. __ H I __ __ __ __
10. __ __ T H

11. N __ C __
12. __ __ T E R
13. L A __ __ __ __
14. __ __ A C K
15. P __ __ C __ L
16. P H __ T __
17. N __ __ B E __
18. __ __ O R
19. M U __ __ C
20. M E __ __ __ __ __

21. L __ V __
22. S T __ __ __ __
23. B __ __ __ N
24. C L __ S __
25. S C __ O __ __
26. H U N __ __ __
27. G R __ __ __
28. B __ O __ __
29. Y __ __ __ O W
30. __ __ Z Z A

31. __ A __ P Y
32. C __ N __ Y
33. B R __ __ __ __ __

34. __ __ __ T E
35. __ __ U G H
36. I M A __ __
37. M I __ __ __ __ __ __
38. G __ __ N T
39. P I __ __ O
40. W E L __ __ __ __

41. __ __ T T __ E
42. C O __ __ __
43. D __ __ S
44. M __ T __ R
45. W __ __ __ E R
46. T O __ __ __
47. __ __ __ N E
48. B __ __ __ N A
49. K __ __ E
50. __ __ I __ G

Appendix C-Demographics Questionnaire

INSTRUCTIONS: READ THE ITEMS BELOW AND CIRCLE THE LETTER THAT BEST DESCRIBES YOU OR WRITE IN THE INFORMATION THAT REFLECTS YOU.

1. Your class standing can be best described as:
    a. Freshman
    b. Sophomore
    c. Junior
    d. Senior

2. Which of the following best describes your racial/cultural identification?
    a. Black American
    b. Black Other
        1b. West Indian (Specify nationality)_____
        2b. African (Specify nationality)_____
        3b. Hispanic (Specify nationality)_____
    c. European American
    d. Latino American (Specify nationality)_____
    e. Other (Specify) _____

3. Sex
    a. male
    b. female

4. Age_____

5. What do you consider your socioeconomic status to be?
    a. Working class
    b. Middle class
    c. Upper middle class
    d. Upper class
    e. Other_____

6. From the list below, please circle the letter indicating your parents' combined income for the last year:
    a. Under $20,000
    b. $20,000-$50,000
    c. $50,001-$75,000
    d. $75,001-$100,000
    e. Over $100,000

7. What is your generational level?

a. First generation (You were born in US, your parents were born in another country) Please list country:_____.
b. Second generation (You and your parents were born in the US, grandparents born in another country).
c. Third generation (You, your parents, and your grandparents were all born in the US)

8. Are you the first in your family to attend college?
    a. Yes
    b. No, parents attended
    c. No, parents and grandparents attended.

9. What is your college cumulative GPA?_____ (give high school GPA if you are a freshman)

10. What is your major?

11. What was your SAT score?_____

12. What was your ACT score?_____

Appendix D-Implicit Theory Inventory

ITI

Please read the following items and indicate your level of agreement.

	Strongly Agree					Strongly Disagree
1. You have a certain amount of intelligence and you can't really do much to change it.	1	2	3	4	5	6
2. Your intelligence is something about you that you can't change very much.	1	2	3	4	5	6
3. You can learn new things, but you can't really change your basic intelligence.	1	2	3	4	5	6

Appendix E – The Disengagement Scale

TDS

Read each item and indicate to what degree it reflects your own thoughts and feelings, using the 7-point scale below. There are no right or wrong answers. Base your responses on your opinion at the present time.

1 Strongly Disagree	2 Disagree	3 Somewhat Disagree	4 Neither Agree Nor Disagree	5 Somewhat Agree	6 Agree	7 Strongly Agree

	SD						SA
1. I feel that standardized achievement tests are fair tests of my abilities.	1	2	3	4	5	6	7
2. In general, I feel that standardized achievement tests are a good measure of my intelligence.	1	2	3	4	5	6	7
3. Most intelligence tests do not really measure what they are supposed to.	1	2	3	4	5	6	7
4. I feel that standardized achievement tests are definitely biased against me.	1	2	3	4	5	6	7
5. I always feel good about myself when I do well on an academic test.	1	2	3	4	5	6	7
6. Being good at academics is an important part of who I am.	1	2	3	4	5	6	7
7. Doing well on intellectual tasks is very important to me.	1	2	3	4	5	6	7
8. I care a great deal about performing well on tests of my intellectual ability.	1	2	3	4	5	6	7
9. It usually doesn't matter to me one way or the other how I do in school.	1	2	3	4	5	6	7
10. I really don't care what tests say about my intelligence.	1	2	3	4	5	6	7
11. No intelligence test will ever change my opinion of how intelligent I am.	1	2	3	4	5	6	7
12. How I do intellectually has little relation to who I really am.	1	2	3	4	5	6	7

Appendix F-Ethnic Identity Scale

EIS

The U.S. is made up of people of various ethnicities. Ethnicity refers to cultural traditions, beliefs, and behaviors that are passed down through generations. Some examples of the ethnicities that people may identify with are Mexican, Cuban, Nicaraguan, Chinese, Taiwanese, Filipino, Jamaican, African American, Haitian, Italian, Irish, and German. In addition, some people may identify with more than one ethnicity. When you are answering the following questions, we'd like you to think about what YOU consider your ethnicity to be.

Please write what you consider to be your ethnicity here _____ and refer to this ethnicity as you answer the questions below. On the following statements indicate the number which corresponds to your attitude. Use the scale below to best represent your answers.

Does not describe me at all	**Describes me a little**	**Describes me well**	**Describes me very well**
1	2	3	4

1. My feelings about my ethnicity are mostly negative.	1	2	3	4
2. I have not participated in any activities that would teach me about my ethnicity.	1	2	3	4
3. I am clear about what my ethnicity means to me.	1	2	3	4
4. I have experienced things that reflect my ethnicity, such as eating food, listening to music, and watching movies.	1	2	3	4
5. I have attended events that have helped me learn more about my ethnicity.	1	2	3	4
6. I have read books/magazines/newspapers or other materials that have taught me about my ethnicity.	1	2	3	4
7. I feel negatively about my ethnicity.	1	2	3	4
8. I have participated in activities that have exposed me to my ethnicity.	1	2	3	4
9. I wish I were of a different ethnicity.	1	2	3	4
10. I am not happy with my ethnicity.	1	2	3	4
11. I have learned about my ethnicity by doing things such as reading (books, magazines, newspapers), searching the internet, or keeping up with current events.	1	2	3	4
12. I understand how I feel about my ethnicity.	1	2	3	4
13. If I could choose, I would prefer to be of a different ethnicity.	1	2	3	4
14. I know what my ethnicity means to me.	1	2	3	4
15. I have participated in activities that have taught me about my ethnicity.	1	2	3	4
16. I dislike my ethnicity.	1	2	3	4
17. I have a clear sense of what my ethnicity means to me.	1	2	3	4

Appendix G-Perceived Stereotype Threat Scale

PST

Please indicate your level of agreement with each of the following statements by circling the number that corresponds most closely with your own feeling.

	Strongly Disagree				Strongly Agree
1. Some people feel that I have less intelligence because of my race.	1	2	3	4	5
2. Tests may be easier for people of my race.	1	2	3	4	5
3. Professors expect me to do poorly on tests because of my race.	1	2	3	4	5
4. In college classes, people of my race often face biased evaluations.	1	2	3	4	5
5. I never worry that people will draw conclusions about my intelligence based on my race.	1	2	3	4	5
6. Tests have been used to discriminate against people from my racial group.	1	2	3	4	5
7. When I take tests, I want to show that people of my race could perform well on it.	1	2	3	4	5
8. A negative opinion exists about how people of my race perform on tests.	1	2	3	4	5

Appendix H-Test Anxiety Inventory

TAI

Please read each statement and select the number which indicates how you feel most of the time.

1. I feel confident and relaxed while taking tests.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
2. While taking final examinations I have an uneasy upset feeling.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
3. Thinking about the grade I may get in a course interferes with my work on tests.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
4. I freeze up on final exams.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
5. During exams I find myself wondering whether I will ever get through school.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
6. The harder I work at taking a test, the more confused I get.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
7. Thoughts of doing poorly interfere with my concentration on tests.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
8. I feel very jittery when taking an important test.
    1. Almost never
    2. Sometimes

3. Often
4. Almost always

9. Even when I am well prepared for a test, I feel very anxious about it.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
10. I start feeling very uneasy just before getting a test paper back.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
11. During tests I feel very tense.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
12. I wish examinations did not bother me so much.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
13. During important exams I am so tense that my stomach gets upset.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
14. I seem to defeat myself when working on important tests.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
15. I feel very panicky when I take an important exam.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
16. If I were to take an important exam, I would worry a great deal about taking it.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
17. During tests I find myself thinking about the consequences of failing.
    1. Almost never
    2. Sometimes

3. Often
4. Almost always

18. I feel my heart beating very fast during important tests.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
19. As soon as an exam is over I try to stop worrying about it, but I just cannot.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always
20. During a course examination I get so nervous that I forget facts I really know.
    1. Almost never
    2. Sometimes
    3. Often
    4. Almost always

Appendix I- State-Trait Anxiety Inventory-State From

STAI Form Y-1

DIRECTIONS: A number of statements which people have used to describe themselves are given below. Read each statement and then blacken in the appropriate circle to the right of the statement to indicate how you feel *right* now, that is, *at this moment*. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

	NOT AT ALL	SOMEWHAT	MODERATELY SO	VERY MUCH SO
1. I feel calm	1	2	3	4
2. I feel secure	1	2	3	4
3. I am tense	1	2	3	4
4. I feel strained	1	2	3	4
5. I feel at ease	1	2	3	4
6. I feel upset	1	2	3	4
7. I am presently worrying about possible misfortunes	1	2	3	4
8. I feel satisfied	1	2	3	4
9. I feel frightened	1	2	3	4
10. I feel comfortable	1	2	3	4
11. I feel self-confident	1	2	3	4
12. I feel nervous	1	2	3	4
13. I am jittery	1	2	3	4
14. I feel indecisive	1	2	3	4
15. I am relaxed	1	2	3	4
16. I feel content	1	2	3	4
17. I am worried	1	2	3	4
18. I feel confused	1	2	3	4
19. I feel steady	1	2	3	4
20. I feel pleasant	1	2	3	4

Appendix J- Rejection Sensitivity-Race-Based

Please read each scenario below and answer the questions that follow.

1.Imagine that you are in class one day, and the professor asks a particularly difficult question. A few people, including yourself, raise their hands to answer the question.

a. How concerned or anxious would you be that the professor would not call on you because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be that the teacher would not call on you because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

2.Imagine that you are in a pharmacy, trying to pick out a few items. While you are looking at the different brands, you notice one of the store clerks glancing your way.

a. How concerned or anxious would you be that the clerk would be suspecting you of stealing a product because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be that the clerk would be suspecting you of stealing a product because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

3. Imagine you have just completed a job interview over the phone. You are in good spirits because the interviewer seemed enthusiastic about your application. Several days later you complete a second interview in person. Your interviewer informs you that they will let you know about their decision soon.

a. How concerned or anxious would you be that the interviewer would not offer you the job because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be that the interviewer would not offer you the job because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

4.It's late at night and you are driving down a country road you're not familiar with. Luckily, there is a 24-hour 7-11 just ahead, so you stop there and head up to the counter to ask the young woman for directions.

a. How concerned or anxious would you be that the woman would be hesitant to talk to you because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be that the woman would not give you directions because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

5.Imagine that a new school counselor is selecting students for a summer scholarship fund that you really want. He has only one scholarship left and you are one of several students that are eligible for the scholarship.

a. How concerned or anxious would you be that the counselor would not select you for the scholarship because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b.How likely do you think it would be that the counselor would not select you for the scholarship because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

6.Imagine that you have just finished shopping, and you are leaving the store carrying several bags. It's closing time and several people are filing out of the store at once. Suddenly the alarm begins to sound, and a security guard comes over to investigate.

a. How concerned or anxious would you be that guard might stop you because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be for the guard to stop you because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

7. Imagine that you are riding the bus one day. The bus is full except for two seats, one of which is next to you. As the bus comes to the stop, you notice a woman getting on the bus.

a. How concerned or anxious would you be that the woman would not sit next to you because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be for the woman to not sit next to you because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

8. Imagine that you are in a restaurant, trying to get the attention of your waitress. A lot of other people are trying to get her attention as well.

a. How concerned or anxious would you be that the waitress would not attend to you because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be for the waitress to not attend to you because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

9. Imagine that you're driving down the street, and there is a police barricade just ahead. The police officers are randomly pulling people over to check driver's licenses and registrations.

a. How concerned or anxious would you be that police officer would pull you over because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be for the police officer to pull you over because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

10. Imagine that it's the second day of your class. The teacher assigned a writing sample yesterday and today the teacher announces that she has finished correcting the papers. You wait for the paper to be returned.

a. How concerned or anxious would you be that you would receive a poor grade because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be for the teacher to give you a poor grade because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

11. Imagine that you are standing in line for the ATM machine, and you notice the woman at the machine glances back while she is getting her money.

a. How concerned or anxious would you be that the woman is glancing back because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it is that the woman is glancing back because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

12. Imagine you're at a pay phone on a street corner. You have to make a call, but you don't have change. You decide to go into a store and ask for change for your bill.

a. How concerned or anxious would you be that the store clerk would not give you change because of your race/ethnicity?

very unconcerned 1 2 3 4 5 6 very concerned

b. How likely do you think it would be that the store clerk would not give you change because of your race/ethnicity?

very unlikely 1 2 3 4 5 6 very likely

Appendix K- Interview Questions

1. What did you think of the test you just took?

2. What do you think was the purpose of the test? What did it measure?

3. How did you feel while taking the test?

4. While taking the test, did you believe you could improve your performance?

5. How effective do you think the test you were given was in determining your ability as a university student?

6. Have there been times when you felt you were negatively stereotyped regarding your intelligence on this campus? Prior to coming to UT? Please describe.

7. Do you believe there are stereotypes for your ethnic group regarding your intelligence?

8. Do you believe there are stereotypes for your gender regarding your intelligence?

9. Have you ever had a time that you were so worried about appearing stereotypical that it interfered with what you were trying to accomplish? Do you think that happened today?

10. Intelligence has sometimes been described as something that you are born with that does not change, meaning that however much you have when you are young stays the same for the rest of your life. Do you agree?

11. Have you ever taken the WAIS? Have you learned about it in any of your classes?

Appendix L - Debriefing Form

I asked you to complete a series of exercises that included arithmetic and problem-solving. I did not tell you the entire purpose of the study because we did not want to influence your performance. However, I will now explain in greater detail.

The exercises that you completed are actually part of an intelligence measure. You completed four subtests of the Wechsler Adult Intelligence Score-IV or the WAIS-IV. We did not reveal this information because sometimes knowing that your intelligence is being evaluated can cause anxiety and disrupt your performance. When your performance may be affected by a negative stereotype about your group, such as that African Americans/Latino Americans/European Americans are/are not intelligent, a phenomenon known as stereotype threat may occur. Stereotype threat is a concept that refers to a negative stereotype influencing a person's performance on a test. We conducted this study to see if stereotype threat occurs in typical testing situations, and if this level of stereotype threat may be reduced.

You were in the malleability/control condition. One goal of the study is to determine whether telling participants that it is possible to improve their performance on the WAIS. Since the purpose of the study was not to evaluate your intelligence, we will not be calculating your IQ. It is our policy not to disclose feedback about your scores since I am a trainee.

You will receive a copy of this debriefing form for your records. We are also providing all participants with a mental health resource list in case you would like to utilize them. I would be happy to answer any questions that you have. If you would like to contact the principal investigator for any reason, Brittany Hall-Clark, M.A., you may email her at brittany.hall@mail.utexas.edu, or by calling 512-471-2170. If you have questions about your rights as a research participant, complaints, concerns, or questions about the research please contact Jody Jensen, Ph.D., Chair, The University of Texas at Austin Institutional Review Board for the Protection of Human Subjects at (512) 232-2685 or the Office of Research Support at (512) 471-8871 or email: orsc@uts.cc.utexas.edu. Thank you for participating in the study.

_____ _____

Participant Date

_____ _____

Person Obtaining Consent Date

_____ _____

Principal Investigator Date

Appendix M- Additional Interview Questions (aka Questions 12 and 13)

1. Do you think you've ever experienced the feelings that you just read about, called stereotype threat? Please explain.

2. Do you think you experienced stereotype threat in today's study?

Appendix N- Resource List

Resource List

The following places are resources that you may contact if you wish to seek help for emotional distress.

UT Counseling and Mental Health Center
Student Services Building, 5th floor
website: www.utexas.edu/student/cmhc/
(512) 471-3515

24-hour Telephone Counseling
(512) 471-CALL(2255)

Manuel Ramirez III, Ph.D
Professor of Clinical Psychology and Licensed Psychologist
(512) 475-7012

Appendix O-Table 1

Descriptives of Sociodemographic Variables

	Overall (*N* = 138)	African American (*N*=29)	Latino American (*N*=57)	European American (*N*=52)
Gender (% female)	52%	59%	60%	39%
SES				
Working Class	20%	24%	33%	4%
Middle Class	41%	45%	47%	31%
Upper Middle Class	30%	21%	18%	50%
Upper Class	8%	7%	2%	15%
Other	1%	4%		
Parent Income				
Under 20k	8%	10%	12%	2%
20k-50k	28%	31%	40%	12%
50k-75k	16%	21%	18%	12%
75k-100k	16%	17%	16%	16%
Over 100k	32%	21%	14%	59%
Did not report	1%			
US Generation				
1st	28%	24%	46%	12%
2nd	20%	3%	28%	20%
3rd	52%	72%	25%	69%
College Generation				
1	28%	38%	46%	2%
2	41%	45%	37%	42%
3	32%	17%	18%	56%
Year in School				
Freshman	59%	59%	56%	62%
Sophomore	20%	10%	25%	20%
Junior	14%	17%	14%	12%
Senior	8%	14%	5%	8%
Mean Age	19.28 (1.51)	19.34 (1.49)	19.21 (1.35)	19.31 (1.71)
Mean GPA	3.45 (.76)	3.18 (1.00)	3.37 (.64)	3.69 (.66)
Mean SAT (N=117)	1599.19 (380.59)	1489.42 (277.77)	1523.91 (401.56)	1739.56 (374.07)
Mean ACT (N=60)	25.62 (4.38)	21.83 (3.24)	24.61 (3.61)	28.36 (3.78)

Appendix P-Table 2

Descriptives of Main Variables

	Overall	African American	Latino American	European American
Manipulation checks:				
Agree with entity?	.20(.40) .80(.42)	.21(.41) .79(.41)	.18 (.38) .88(.38)	.23(.43) .73(.45)
Believe improve?				
Wais:				
BD	45.50(12.09)	37.86(12.34)	46.21(10.99)	48.98(11.42)
DS	29.63(4.98)	27.79(4.79)	29.00(4.64)	31.35(5.02)
MR	20.49(3.34)	19.69(3.60)	20.40(3.53)	21.04(2.90)
AR	15.01(2.90)	13.03(2.80)	14.75(2.85)	16.38(2.26)
Implicit Theories	4.61 (1.23)	4.97(.99)	4.77(1.13)	4.24(1.37)
Ethnic identity-exploration	2.89 (.85)	3.28(.72)	3.08(.72)	2.47(.88)
Ethnic identity-Affirmation	3.82(.39)	3.68(.50)	3.82(.37)	3.90(.33)
Ethnic identity-Resolution	3.18(.79)	3.29(.70)	3.39(.70)	2.88(.86)
Perceived Stereotype Threat	2.76 (.92)	3.57(.73)	3.06(.72)	1.98(.54)
Stereotype Vulnerability	188.68(189.59)	351.66(242.49)	195.93(169.47)	89.85(87.29)
Pre-Test Test Anxiety	36.22(10.90)	39.83(11.63)	37.39(12.12)	32.92(7.98)
Post-Test Test Anxiety	36.31(12.76)	40.93(13.20)	38.25(14.40)	31.62(8.67)
Pre-Test State Anxiety	31.97(9.24)	31.38(9.64)	33.35(9.59)	30.79(8.58)
Post-test State Anxiety	32.83(10.59)	33.24(11.17)	35.26(11.67)	29.92(8.24)
Discounting	4.13(1.20)	4.85(1.09)	3.89(1.10)	3.98(1.22)
Devaluing	2.13(.96)	2.42(1.18)	1.86(.65)	2.26(1.06)
Disengagement	1.25(1.25)	4.11(1.34)	3.52(1.21)	3.76(1.21)

Appendix Q-Table 3

Intercorrelations of Main Variables

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
1. Block Design	-										
2. Digit Span	0[ł]	-									
3. Matrix Reasoning	.33[***]	.31[***]	-								
4. Arithmetic	.37[***]	.47[***]	.40[***]	-							
5. Implicit Theory of Intelligence	-.13	-.22[*]	-.08	-.26[**]	-						
6. Perceived Stereotype Threat	-.19*	-.21[*]	-.10	- .35[***]	.15[ł]	-					
7. Stereotype Vulnerability	-.34[***]	-.20[*]	-.15[ł]	-.22[**]	.08	.61[***]	-				
8. Ethnic Identity Exploration Affirmation Resolution	-.25[**] -.04 -.21[*]	-.19[*] -.01 -.03	.11 .08 .17[*]	-.17[*] .10 -.13	.32[***] .01 .12	.37[***] -.18[*] .16[ł]	.33[***] -.14[ł] .09	-			
9. Disengagement Discounting Devaluing Disengaging	-.18[*] -.07 .01	-.15 -.06 -.08	-.17[*] -.06 -.08	-.22[**] -.06 -.03	.25[**] .01 .21[*]	.32[***] -.08 -.07	.30[***] .08 .03	.20[*] -.10 -.02 -.10 .03 -.30[***] .09 .08 -.03	-		
10. Test Anxiety Pre-Test Post-Test	-.11 -.13	-.14[ł] -.18[*]	-.14[ł] -.18[*]	-.14 -.21[**]	.09 .12	.33[***] .37[***]	.34[***] .37[***]	.08 -.04 -.08 .12 -.12 -.08	.27[**] .04 .04 .31[***] .08 .01	-	
11. State Anxiety Pre-Test Post-Test	-.08 -.07	-.09 -.20[*]	-.09 -.20[*]	-.15[ł] -.18[*]	.01 .05	.18[*] .28[***]	.30[***] .32[***]	-.12 -.11 -.17[*] -.04 -.22[**] -.17[*]	.21[**] .04 -.10 .30[***] .09 -.11	.42[***] .42[***] .45[***].52[***]	-

[ł] = .10
[*] = .05
[**] = .01
[***]= .001

Appendix R-Table 4

Chi Square Analyses and ANOVA

	χ^2	*df*	*p*
Year in School	.79	3	.85
Race	.13	2	.94
Gender	.26	1	.61
SES	2.59	4	.63
Income	8.18	4	.09
US Generation	2.03	3	.57
College Generation	1.59	2	.45
Manipulation Checks			
Believe could improve while taking test?	.79	1	.37
Believe born with intelligence and does not change?	.18	1	.67

	F	*df*	*p*
Age	.56	1, 137	.45
GPA	.80	1,135	.37
SAT	3.56	1,116	.06
ACT	.06	1, 59	.81
Manipulation Check:			
ITI	.57	1,135	.45

Appendix S-Table 5

Moderation Analyses

	Block Design			Digit Span			Matrix Reasoning			Arithmetic		
	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE
Step 1:	.08			.06			.05			.13		
PST		-1.59	1.19		-.86-	.49		-.54	.34		-.94**	.32
EIS-exp		-2.02	1.57		-.96	.62		.35	.36		-.11	.79
EIS-aff		-.64	2.80		-.17	1.39		.21	1.00		.61	.45
EIS-res		-1.83	1.45		.47	.63		.64	.40		-.25	.36
Step 2:	.03			.01			.01			.01		
PST		-.76	1.35		-.88	.54		-.60	.39		-.94**	.32
EIS-exp		-2.78 -	1.57		-.89	.67		.43	.40		-.11	.38
EIS-aff		1.13	3.92		-.53	.1.73		-.09	1.20		.61	.97
EIS-res		-2.12	1.51		.46	.67		.65	.45		-.25	.36
PSTxexp		-.3.18*	1.37		.03	.64		.20	.64		-.13	.36
PSTxaff		-.32	4.48		.77	1.85		.38	1.58		-.52	1.07
PSTxres		.29	1.31		.44	.68		.23	.59		.18	.32

- = marginal * = .05 or less ** = or less .01

Appendix T-Table 6

Interview Question 2: *What did you think was the purpose? What did it measure?*

Purpose of Test	Overall	AA	EA	LA
Ethnic Differences	2%	0%	0%	4%
Under Pressure	9%	10%	12%	7%
Adaptability	6%	0%	6%	12%
Ability	59%	48%	66%	55%
Speed	23%	34%	21%	20%
Intelligence	10%	3%	10%	13%
Knowledge	3%	0%	0%	4%

Appendix U-Table 7

Interview Question 3: *How did you feel while taking the exam?*

Emotional Experience	Overall	AA	EA	LA
Negative feelings	38%	38%	45%	32%
Generally calm	49%	59%	50%	45%
Apathetic	2%	0%	0%	4%
Stupid/slow	6%	3%	6%	7%
Ambiguous	6%	3%	12%	5%
Related to difficulty	25%	20%	6%	18%

Appendix V-Table 8

Interview Question Four: *While taking the test, did you believe you could improve your performance?*

Response	Overall	AA	EA	LA
No (entity)	19%	21%	27%	11%
Maybe	5%	7%	6%	5%
Qualified yes	8%	7%	4%	11%
Yes (incremental)	65%	66%	60%	71%

Appendix W-Table 9

Interview Question Five: *How effective do you think the test you were given was in determining your ability as a university student?*

Response	Overall	AA	EA	LA
Not effective	53%	55%	54%	54%
Qualified yes	29%	24%	29%	32%
Effective	15%	21%	17%	13%

Appendix X-Table 10

Interview Question Six: *Have there been times when you felt you were negatively stereotyped regarding your intelligence on this campus? Prior to coming to UT?*

Discrimination	AA	EA	LA
At UT	38%	27%	34%
Prior to UT	31%	35%	30%

Appendix Y-Table 11

Interview Question Seven: *Do you believe there are stereotypes for your ethnic group regarding your intelligence?*

Awareness of Racial/Ethnic Stereotypes	AA	EA	LA
Yes	97%	54%	92%
No	3%	40%	9%

Appendix Z-Table 12

Interview Question Eight: *Do you believe there are stereotypes for your gender regarding your intelligence?*

Response	AA	EA	LA	Female	Male
Yes	72%	67%	59%	62%	53%
No	28%	33%	41%	38%	48%

Appendix AA-Table 13

Interview Question Nine: *Have you ever had a time that you were so worried about appearing stereotypical that it interfered with what you were trying to accomplish? Do you think that happened today?*

Response	Overall	AA	EA	LA
Yes, during WAIS-IV	4%	11%	2%	4%
Yes, not during WAIS-IV	10%	11%	10%	11%
No	85%	76%	89%	86%

Appendix AB-Table 14

Interview Question Ten: *Intelligence has sometimes been described as something that you are born with that does not change, meaning that however much you have when you are young stays the same for the rest of your life. Do you agree?*

Response	Overall	AA	EA	LA
Yes	11%	14%	10%	11%
Somewhat	14%	14%	21%	7%
No	75%	72%	69%	82%

Appendix AC-Table 15

Interview Question Eleven: *Have you ever taken the WAIS? Have you been exposed to it in any class?*

Response	Overall	AA	EA	LA
Never heard of	75%	69%	77%	77%
Taken	5%	7%	6%	4%
Heard about	18%	21%	17%	20%

Appendix AD-Table 16

Interview Question Twelve: *Do you think you've ever experienced the feelings that you just read about, called stereotype threat? Please explain.*

Response	Overall	AA	EA	LA
Yes	32%	45%	81%	63%
No	65%	48%	17%	36%

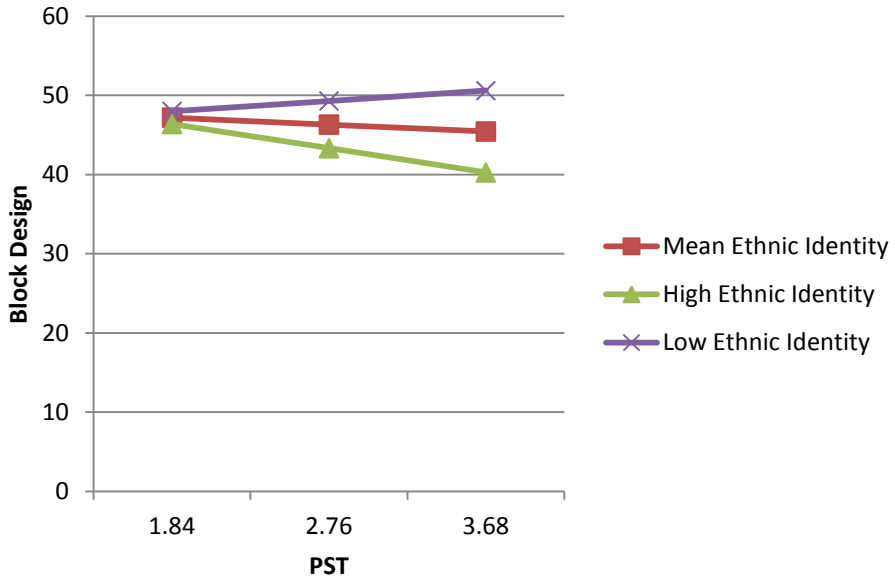
Appendix AE-Table 17

Interview Question Thirteen: *Do you think you experienced stereotype threat in today's study?*

Response	Overall	AA	LA	EA
Yes	5%	10%	0%	5%
No	95%	90%	100%	95%

Appendix AF-Figure

PST x Ethnic Identity-Exploration: Block Design



REFERENCES

Armenta, B.E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology, 16,* 94-98.

Aronson, J. (2002). Stereotype threat: Contending and coping with unnerving expectations. In J. Aronson, *Improving academic achievement: Impact of psychological factors on education* (pp. 279-301). San Diego, CA, US: Academic Press.

Aronson, J., & Fried, C. (1997). *Belief in the malleability of intellectual ability: An intervention to increase the performance of African-American college students.* Unpublished manuscript, University of Texas, Austin, Texas.

Aronson, J., Fried, C.B., & Good, C (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38,* 113-125.

Aronson, J. & Inzlicht, M. (2004). The ups and downs of attributional ambiguity: Stereotype vulnerability and the academic self-knowledge of African American college students. *Psychological Science, 15,* 829-836.

Aronson, J., Lustina, M. J., Good, C. & Keough, K. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35,* 29-46.

Aronson, J., Quinn, D.M., & Spencer, S.J. (1998). Stereotype threat and the

academic underperformance of minorities and women. In J.K. Swim & C. Stangor (Eds.), *Prejudice: The Target's Perspective* (pp. 83-103). San Diego, CA: Academic Press.

Aronson, J., & Salinas, M.F. (1997). *Stereotype threat, attributional, ambiguity, and Latino underperformance.* Unpublished manuscript, University of Texas, Texas, Austin.

Aronson, J. & Steele, C. M. (2005). Stereotypes and the fragility of academic competence, motivation, and self-concept. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of Competence and Motivation* (pp. 436-456). New York: Guilford Publications.

Aronson, J., Tichy, J.C., & Croteau (1997). *Conceptions of ability and stereotype threat.* Unpublished manuscript, University of Texas, Austin, Texas.

Arredondo, P., & Perez, P. (2003). Counseling paradigms and Latina/o Americans: Contemporary considerations. In F. Harper & J. McFadden (Eds.), *Culture and Counseling: New Approaches* (pp. 115-132). Boston: Allyn and Bacon.

Awad, G. H. (2007). The role of racial identity, academic self-concept, and self-esteem in the prediction of academic outcomes for African American students. *Journal of Black Psychology, 33*, 188-207.

Barnes, L. L. B., Harp, D. & Jung, W. S. (2002). Reliability generalization of scores on the State-Trait Anxiety Inventory. *Educational and Psychological Measurement, 62,* 603-618.

Beilock, S.L., Jellison, W.A., Rydell, R.J., McConnell, AR., & Carr, T.H. (2006). On the causal mechanisms of stereotype threat: can skills that don't rely heavily on

working memory still be threatened? *Personality and Social Psychology Bulletin, 32,* 1059-1071.

Berry, J. (2005). Living successfully in two cultures. *Journal of Intercultural Processes, 29,* 697-712.

Blackwell, L.S., Trzesniewski, K.H., & Sorich Dweck, C. (2007). Implicit theories of Intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78,* 246-263.

Brody, N. (2000). History of theories and measurements of intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 3-15). New York: Cambridge University Press.

Brown, R. P. & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology, 91,* 979-985.

Burns, K. C. & Isbell, L. M. (2007). Promoting malleability is not one size fits all: Priming implicit theories of intelligence as a function of self-theories. *Self and Identity, 6,* 51-63.

Buss, W. G. (1996). Intelligence testing and judicial policy making for special education. *Psychology, Public Policy, and Law, 2,* 584-602.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? *Psychological Science, 16,* 572-578.

Ceci, S J. (1996). General intelligence and life success: An introduction to the special theme. *Psychology, Public Policy, and Law,* 2, 403-417.

Cohen, G. L. & Garcia, J. (2005). "I am us": Negative stereotypes as collective threats. *Journal of Personality and Social Psychology, 89,* 566-582.

Cohen, G.L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*, 1307-1310.

Cohen, G. L. & Sherman, D. K. (2005). Stereotype threat and the social and scientific contexts of the race achievement gap. *American Psychologist, 60.* (Date of access: 6/10/2005):http://web31.epnet.com.content.lib.utexas.edu:2048/citation.asp?tb=1&_ug=sid+50200653

Cokley, K. (2006). The Impact of Racialized Schools and Racist (Mis)Education on African American Students' Academic Identity. In Constantine, M. G., & Sue, D. W. (Eds.), *Addressing racism: Facilitating cultural competence in mental health and educational settings*. (pp. 127-144). Hoboken, NJ: John Wiley & Sons Inc.

Croizet, J., Després, G., Gauzins, M., Huguet, P., Leyens, J. & Méot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30,* 721-731.

Cullen, M. J., Hardison, C. & Sackett, P. R. (2004). Using SAT-Grade and Ability-Job Performance Relationships to Test Predictions Derived From Stereotype Threat Theory. *Journal of Applied Psychology, 89,* 220-230.

Danaher, K. & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38,* 1639-1655.

Davis III, C., Aronson, J. & Salinas, M. (2006). Shades of threat: Racial identity as a moderator of stereotype threat. *Journal of Black Psychology, 32,* 399-417.

Davis, C. & Simmons, C. (2009). Stereotype threat: A review, critique, and implications. In Neville, H.A., Tynes, B. M., and Utsey, S. O. (Eds.), *Handbook of African American Psychology* (pp. 211-222). Thousand Oaks, CA, US: Sage Publications, Inc.

Dweck, C. S., Chiu, C. & Hong, C. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6,* 267-285.

Dweck, C. S. & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256-273.

Eysenck, H. J. (1984). The effect of race on human abilities and mental test scores. In C. R. Reynolds and R. T. Brown (Eds)., *Perspectives on Bias in Mental Testing* (pp. 249-291). New York: Plenum.

Fagan, J.F., & Holland, C.R. (2009). Culture-fair prediction of academic achievement. *Intelligence*, *37*, 62-67.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*, 29-51.

Giorgi, A. P. & Giorgi, B. (2008). Phenomenological psychology. In C. Willig and W. Stainton-Rogers (Eds). *The SAGE Handbook of Qualitative Research in Psychology* (pp. 165-178). SAGE Publications: Los Angeles.

Goldstein, H. W., Scherbaum, C. A. & Yusko, K. P. (2010). Revisiting g: Intelligence, adverse impact, and personnel selection. I*:* J. L Outtz,. (Ed.) *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 95-134.) New York: Routledge/Taylor & Francis Group.

Gonzales, P. M., Blanton, H. & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin, 28,* 659-670.

Gonzales, R. R. & Roll, S. (1985). Relationship between acculturation, cognitive style and intelligence: A cross-sectional study. *Journal of Cross-Cultural Performance, 16,* 190-205.

Good, C., Aronson, J., & Harder, J.A. (2007). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29,* 17-28.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Development Psychology, 24,* 645-662.

Gottfredson, L., &Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology, 50,* 183-195.

Guthrie, R. (1976). *Even the rat was white*. New York: Harper and Row.

Guyll, M., Madon, M., Prieto, L. & Scherr, K. C. (2010). The potential roles of self-fulfiling prophecies, stigma consciousness, and stereotype threat in linking Latino/a identity ethnicity and educational outcomes. *Journal of Social Issues, 66,* 113-130.

Harrison, L. A., Stevens, C. M., Monty, A. N., & Coakley, C. A. (2006). The consequences of stereotype threat on the academic performance of White and non-White lower income college students. *Social Psychology of Education, 9,* 341-357.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychology, 47,* 1083-1101.

Herrnstein, R. J. & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. *New York: The Free Press.*

Hill, N. E. & Torres, K. (2010). Negotiating the American dream: The paradox of aspirations and achievement among Latino students and engagement between their families and schools. *Journal of Social Issues, 66,* 95-112.

Howell, D. C. (2002). *Statistical methods for psychology (5th edition).* Duxbury Thomson Learning: Pacific Grove.

Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99,* 545-560.

Hunt, E. & Carlson, J. (2007) Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science, 2,* 194-213.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39,* 1-123.

Johns, M. Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16,* 175-179.

Jordan, A. H. & Lovett, B. J. (2007). Stereotype threat and test performance: A primer for school psychologists. *Journal of School Psychology, 45,* 45-59.

Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, Race, Residence, Region and Education differences on the II WAIS-R subtests. *Journal of Clinical Psychology, 44,* 231-248.

Kaufman, McLean, & Reynolds (1991). Analysis of WAIS-R factor patterns by sex and Race. *Journal of Clinical Psychology*, 47, 548-557.

Kenny, D. A. (2008). Mediation. Retrieved from http://davidakenny.net/cm/mediate.htm#BK

Kim, B. S. K. & Abreu, J. M. (2001). Acculturation measurement: Theory, current measurements, and future directions. In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, &C. Alexander (Eds.), *Handbook of multicultural counseling* (2nd ed., pp. 394-204). Thousand Oaks, CA: Sage Publications.

Krendl, A.C., Richeson, J.A., Kelley, W.M., & Heatherton, T.F. (2008). The negative consequences of threat: An fMRI investigation of the neural mechanisms underlying women's underperformance in math. *Psychological Science, 19*, 168-175.

Landrine, H., & Klonoff, E. A. (1996). *African American acculturation: Deconstructing race and reviving culture*. Thousand Oaks, CA: Sage.

Levy, S.R., Stroessner, S.J., & Dweck, C.S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74*, 1421-1436.

Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 176-193). New York: Cambridge University Press.

Lohman, D. F. (1997.) The history of intelligence testing in context: The impact of personal, religious, and scientific beliefs on the development of theories and tests of human abilities. In R. F Dillon (Ed.), *Handbook on testing* (pp. 82-106). Westport, CT, US: Greenwood Press/Greenwood Publishing Group.

Lowery, B. S. & Wout, D. A. (2010). When inequality matters: The effect of inequality frames on academic engagement. *Journal of Personality and Social Psychology, 98,* 956-966.

Mackintosh, N. J. (1998) IQ and human intelligence. New York: Oxford University Press

Major & Schmader (1998). Coping with stigma through psychological disengagement. In J. K. Swim & C. Stangor (Eds.) *Prejudice: The target's perspective* (pp. 219-241). San Diego: Academic Press, Inc.

Martens, A., Johns, M., Greenburg, J, & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42,* 236-243.

Marx, D. M. & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology, 44,* 645-657.

Mayer, D. M. & Hanges, P. J. (2003). Understanding the stereotype threat effect with "culture-free" tests: An examination of its mediators and measurement. *Human Performance, 16,* 207-230.

McFarland, L. A., Lev-Arey, D. M., & Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Human Performance, 16,* 181-205.

McGrew, K. S. & Flanagan, D. P. (1998). The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment. Needham Heights: Allyn & Bacon.

McKay, P. F., Doverspike, D., Bowen-Hilton, D., & Martin, Q. D. (2002). Stereotype threat effects on the Raven Advanced Progressive Matrices scores of African Americans. *Journal of Applied Social Psychology, 32,* 767-787.

McKown, C. & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development, 74,* 498-515.

Mendoza-Denton, R., Downey, G., Purdie, V. J., Davis, A. & Pietrzak, J. (2002). Sensitivity to status-based rejection: Implications for African American students' college experience. *Journal of Personality and Social Psychology,* 83, 896–918.

Miller-Jones, D. (1989). Culture and testing. *American Psychologist, 44,* 360-366.

Murdoch, S. (2007). IQ: The smart history of a failed idea. New Jersey: John Wiley & Sons, Inc.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J. & Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist, 51,* 77-101.

Nguyen, H. D. & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93* 1314-1334.

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count.* New York: Norton, W. W. & Company, Inc.

Nussbaum, A. D. & Dweck, C. S. (2008). Defensiveness versus remediation: Self-theories and modes of self-esteem maintenance. *Personality and Social Psychology Bulletin, 34,* 599-612.

Oei, T. P, Evans. L., & Crook, G. M. (1990). Utility and validity of the STAI with anxiety disorder patients. *British Journal of Clinical Psychology, 29,* 429-432.

Olmedo, E. L. (1981). Testing linguistic minorities. *American Psychologist, 36,* 1078-1085.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology, 26,* 291-310.

Osborne, J. W. (2007). Linking stereotype threat and anxiety. *Educational Psychology, 27,* 135-154.

Phinney, J. S. (1996). When we talk about American ethnic groups, what do we mean? *American Psychologist, 51,* 918-927.

Ployhart, R. E., Ziegert, J. C. & McFarland, L. A. (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactions research. *Human Performance, 16,* 231-259.

Quinn, D.M., & Spencer, S.J. 2001. The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57,* 55-71.

Ramirez, M. & Gonzalez, A. (1972). Mexican Americans and intelligence testing. In M. M. Mangold (Ed.), *La causa chicana: The movement for justice.* New York: Family Service Association of America.

Reeve, C.L. & Bonaccio, S. (2008). Does test anxiety induce measurement bias in cognitive ability tests? *Intelligence, 36,* 526-538.

Ryan, K.E., & Ryan, A. M. (2005). Psychological processes underlying stereotype threat and standardized math test performance. *Educational Psychologist, 40,* 53-63.

Rydell, R.J., McConnell, A.L., & Beilock, S.L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality & Social Psychology, 96,* 949-966.

Rosenthal, H.E.S., & Crisp, R.J. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin, 32,* 501-511.

Rosenthal, H.E.S., Crisp, R.J., & Suen, M.W. (2007). Improving performance expectations in stereotypic domains: Task relevance and the reduction of stereotype threat. *European Journal of Social Psychology, 37,* 586-597.

Rushton, J.P. (2005). Ethnic nationalism, evolutionary psychology, and genetic similarity theory. *Nations and Nationalism*, *11*, 489-507.

Rushton, J.P. & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11,* 235-294.

Sackett, P. R., Hardison, C. M. & Cullen, M. J. (2004a). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist, 59,* 7-13.

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004b). On the value of correcting mischaracterizations of stereotype threat research. *American Psychologist, 59,* 48-49.

Sackett, P.R., Schmitt, N., Ellingson, J. E. & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist, 56,* 302-318.

Sanchez, G. I. (1932). Group differences and Spanish-speaking children—a critical review. *Journal of Applied Psychology*, *16*, 549-558.

Sappington, A. & Grizzard, R. (1975). Self-discrimination responses in black school children. *Journal of Personality and Social Psychology, 31*, 224-231.

Sarason, I.G. (1978). The Test Anxiety Scale: Concept and research. Test Anxiety: Theory, Research, and Applications. In C.D. Spielberger & I.G. Sarason (Eds.), Stress and anxiety (Vol. 5). Washington, D.C.: Hemisphere Publishing Corp.,

Sawyer, T. P. & Hollis-Sawyer, L. A. (2005). Predicting stereotype threat, test anxiety, and cognitive ability exam performance: An examination of three models. *International Journal of Testing, 5*, 225-246.

Schmader, T. & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 440-452.

Sellers, R. M., Chavous, T. M., & Cooke, D. Y. (1998). Racial ideology and racial centrality as predictors of African American college students' academic performance. *Journal of Black Psychology, 24*, 8-27.

Shih, M. Pittinsky, T. L. & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.

Smedley, A. & Smedley B. D. (2005). Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. *American Psychologist, 60*, 16-26.

Snyderman, M. & Rothman, S. (1987). Survey of expect opinion on intelligence and aptitude testing. *American Psychologist, 42*, 137-144.

Spencer, B. & Castano, E. (2007). Social class is dead. Long live social class! Stereotype threat among low socioeconomic status individuals. *Social Justice Research, 20*, 418-432.

Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35,* 4-28.

Spielberger, C. D. (1980). Test Anxiety Inventory. *Preliminary professional manual*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D. (1983). Manual for the State-Trait Anxiety Inventory (STAI). Palo Alto, CA: Consulting Psychologists Press.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613-629.

Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-811.

Steele, C. M. & Aronson, J. A. (2004). Stereotype threat does not live by Steele & Aronson (1995) alone. *American Psychologist, 59,* 47-48.

Steele, C. M. & Davies, P. G. (2003). Stereotype threat and employment testing: A commentary. *Human Performance, 16,* 311-326.

Sternberg, R. J. (2000). The concept of intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 3-15). New York: Cambridge University Press.

Stricker, L. J. & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology, 34,* 665-693.

Stricker, L. J. & Ward, W. C. (2008). Stereotype threat in applied settings re-examined: A reply. *Journal of Applied Psychology, 38,* 1656-1663.

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., and Esquilin, M. (2007). Racial Microaggressions in Everyday Life. *American Psychologist, 62,* 271-286.

Terrell, F., Terrell, S. & Taylor, J. (1981). Effects of race of examiner and cultural mistrust on the WAIS performance of Black students. *Journal of Consulting and Clinical Psychology,* 49, 750-751.

Thorndike, R. M. (1997). The early history of intelligence testing. In Flanagan, D. P., Genshaft, J. L. & Harrison, P. L. (Eds.) *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3-16). New York: Guilford Press.

Umaña-Taylor, A. J. Yazedjian, A., & Bámaca-Gómez, M. (2004). Developing the Ethnic Identity Scale using Eriksonian and social identity perspectives. *Identity: An International Journal of Theory and Research, 4,* 9-38.

Valencia, R., & Suzuki, L.A. (Eds.). (2001) *Intelligence testing and minority students: Foundations, performance factors, and assessment issues*. Thousand oaks, CA: Sage.

Van Laar, C., Levin, S., & Sinclair, S. (2008). Social identity and personal identity stereotype threat: The case of affirmative action. *Basic and Applied Social Psychology, 30,* 295-310.

Vincent, K. R. (1991). Black/White IQ differences: Does age make the difference? *Journal of Clinical Psychology, 47,* 266-270.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-Fourth Edition: Technical and interpretative manual.* San Antonio: Pearson.

Whaley, A.L. (1998). Issues of validity in empirical tests of stereotype threat theory. *Journal of Psychology, 132,* 317-327.

Wicherts, J. M., Dolan, C. V. & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology, 89,* 696-716.

Williams, J. (2005). Forewarning: A tool to disrupt stereotype threat effects. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 65(8-A),* pp. 2901.

Williams, W. M. 3(1996) Consequences on how we define and assess intelligence. *Psychology, Public Policy, and Law, 2,* 506-535.

Wout, D. A., Shih, M.J., Jackson, J.S. and Sellers, R. M. (2009). Targets as perceivers: How people determine when they will be negatively stereotyped. *Journal of Personality and Social Psychology, 96,* 349-362.

Yeung & Pfeiffer, (2009). The Black-White test score gap and early home environment. *Social Science Research, 38,* 412-437.

Yopyk, D.J.A. & Prentice, D.A. (2005). Am I an athlete or a student? Identity salience and stereotype threat in Student-Athletes. *Basic and Applied Social Psychology, 27,* 329-336.

Vita

Brittany Nicole Hall-Clark graduated summa cum laude from the University of Southern California in 2005 with a double major in Psychology and English and a minor in Spanish. She entered graduate school under the mentorship of Dr. Manuel Ramirez, where she served as the manager of the Laboratory of Multicultural Processes and Mental Health. She also worked on the research teams of Dr. Kevin Cokley and Dr. Germine Awad. She has presented at numerous conferences and has eight publications to date. After graduating from the University of Texas at Austin, she will be completing a two-year post-doctoral fellowship with the STRONG STAR Multidisciplinary PTSD Research Consortium where she will work on Dr. Edna Foa's Prolonged Exposure study. Her professional interests include the development of culturally sensitive treatment and assessment, acculturative stress, racial identity, the impact of trauma on survivors and their families, and relaxation.

Permanent email: DrBHallClark@gmail.com

This dissertation was typed by Brittany Nicole Hall-Clark.